

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 23:15:18 ; Search time 1854.01 seconds  
(without alignments)  
11362.978 Million cell updates/sec

Title: US-09-645-192-1  
Perfect score: 1362  
Sequence: 1 atgaagatattcaatgtta.....atccactaccacatcatga 1362

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_ba3:  
4: gb\_in1:  
5: gb\_in2:  
6: gb\_in3:  
7: gb\_om:  
8: gb\_ov:  
9: gb\_pa1:  
10: gb\_pa2:  
11: gb\_ph:  
12: gb\_pi1:  
13: gb\_pi2:  
14: gb\_pi3:  
15: gb\_pi4:  
16: gb\_ba1:  
17: gb\_ba2:  
18: gb\_fun:  
19: gb\_hqo\_hum:  
20: gb\_hqo\_inv:  
21: gb\_hqo\_rod:  
22: gb\_hqo\_hum1:  
23: gb\_hqo\_hum2:  
24: gb\_hqo\_hum3:  
25: gb\_hqo\_hum4:  
26: gb\_hqo\_hum5:  
27: gb\_hqo\_hum6:  
28: gb\_hqo\_hum7:  
29: gb\_hqo\_hum8:  
30: gb\_hqo\_inv1:  
31: gb\_hqo\_inv2:  
32: gb\_hqo\_other:  
33: gb\_hqo\_rod:  
34: gb\_hum1:  
35: gb\_hum2:  
36: gb\_hum3:  
37: gb\_hum4:  
38: gb\_hum5:  
39: gb\_hum6:  
40: gb\_hum7:  
41: gb\_in:  
42: gb\_om:  
43: gb\_or:

44: gb\_ov:  
45: gb\_pa1:  
46: gb\_pa2:  
47: gb\_ph:  
48: gb\_pi1:  
49: gb\_pi2:  
50: gb\_pi3:  
51: gb\_pi4:  
52: gb\_ba1:  
53: gb\_ba2:  
54: gb\_fun:  
55: gb\_hqo\_hum:  
56: gb\_hqo\_inv:  
57: gb\_hqo\_rod:  
58: gb\_hqo\_hum1:  
59: gb\_hqo\_hum2:  
60: gb\_hqo\_hum3:  
61: gb\_hqo\_hum4:  
62: gb\_hqo\_hum5:  
63: gb\_hqo\_hum6:  
64: gb\_hqo\_hum7:  
65: gb\_hqo\_hum8:  
66: gb\_hqo\_inv1:  
67: gb\_hqo\_inv2:  
68: gb\_hqo\_other:  
69: gb\_hqo\_rod:  
70: gb\_hum1:  
71: gb\_hum2:  
72: gb\_hum3:  
73: gb\_hum4:  
74: gb\_hum5:  
75: gb\_hum6:  
76: gb\_hum7:  
77: gb\_hum8:  
78: gb\_in1:  
79: gb\_in2:  
80: gb\_in3:  
81: gb\_om:  
82: gb\_ov:  
83: gb\_pa1:  
84: gb\_pa2:  
85: gb\_ph:  
86: gb\_pi1:  
87: gb\_pi2:  
88: gb\_pi3:  
89: gb\_pi4:  
90: gb\_ba1:  
91: gb\_ba2:  
92: gb\_fun:  
93: gb\_hqo\_hum:  
94: gb\_hqo\_inv:  
95: gb\_hqo\_rod:  
96: gb\_hqo\_hum1:  
97: gb\_hqo\_hum2:  
98: gb\_hqo\_hum3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	10	AX087935 Sequence
2	1362	100.0	3435	89	AF132035 Homo sapi
3	207.2	15.2	3508	94	AB037596 Mus muscu
4	194.8	14.3	108873	58	AF318573 Bovine he
5	194.4	14.3	1206	94	AB037597 Mus muscu
6	191.8	14.1	1317	10	AX087948 Sequence
7	191.8	14.1	1317	88	AF038650 Homo sapi
8	191.8	14.1	2128	88	AF102542 Homo sapi

9	191.8	14.1	2216	9	AX045249	AX045249 Sequence
10	191.8	14.1	181983	57	AC022480	AC022480 Homo sapi
11	191.6	14.1	2017	58	AF231105	AF231105 Bovine he
12	191.2	14.0	2393	94	MM068182	MM068182 Mus musculu
13	187.4	13.8	87636	90	AL138039	AL138039 Human DNA
14	187.4	13.8	176544	78	AL138877	AL138877 Homo sapi
15	187.4	13.8	196788	66	AC021626	AC021626 Homo sapi
16	185.2	13.6	1203	10	AX087950	AX087950 Sequence
17	185.2	13.6	1807	10	I16989	I16989 Sequence 13
18	185.2	13.6	1807	10	I195029	I195029 Sequence 13
19	185.2	13.6	1807	93	HSBNGST	HSBNGST Homo sapiens m
20	185.2	13.6	1807	97	HOMNKGAT	HOMNKGAT Human I bet
21	184.2	13.5	157102	75	AC076970	AC076970 Homo sapi
22	183.6	13.5	155359	90	AL358777	AL358777 Human DNA
23	183.6	13.5	196788	66	AC021626	AC021626 Homo sapi
24	181.2	13.3	252504	75	AC073815	AC073815 Mus muscu
25	180.8	13.3	2217	9	AX045251	AX045251 Sequence
26	180.8	13.3	2217	9	AX045253	AX045253 Sequence
27	179.6	13.2	2033	94	MM019365	MM019365 Mus musculu
28	179.2	12.9	159930	78	AL138757	AL138757 Homo sapi
29	175.2	12.9	183858	90	AL161626	AL161626 Human DNA
30	174.4	12.8	5010	95	S79797	S79797 enzymatic g
31	174	12.8	2060	93	HSIGN1	HSIGN1 Human beta-
32	173.2	12.7	2023	94	DB7332	DB7332 Mus musculu
33	173.2	12.7	2047	94	DB7332	DB7332 Mus musculu
34	172.8	12.7	4179	58	BHY4RGF	BHY4RGF Bovine hearp
35	172	12.6	1287	10	AX087946	AX087946 Sequence
36	172	12.6	2105	10	I14262	I14262 Sequence 3
37	172	12.6	2105	10	I61440	I61440 Sequence 3
38	172	12.6	2105	10	I72548	I72548 Sequence 3
39	172	12.6	2110	97	HUMGCG	HUMGCG Human beta-
40	172	12.6	2204	97	HUMGCG	HUMGCG Human beta-
41	164.8	12.1	2435	7	BT041330	BT041330 Bos taurus
42	158.4	11.6	94171	93	HS1153D9	HS1153D9 Human DNA
43	121.2	8.9	378	10	I16983	I16983 Sequence 1
44	121.2	8.9	378	10	I195023	I195023 Sequence 1
45	113.4	8.3	63853	77	AC087663	AC087663 Homo sapi

## ALIGNMENTS

RESULT 1  
 LOCUS AX087935 1362 bp DNA PAT 17-MAR-2001  
 DEFINITION Sequence 1 from Patent WO0114535.  
 ACCESSION AX087935  
 VERSION AX087935.1 GI:13396913

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## SOURCE

## CDS

human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 1362)  
 Udp-n-acetylglucosamine: galactose-4-epimerase [EC 5.1.3.1]  
 -acetylglucosamine-4-epimerase [EC 5.1.3.1]  
 Patent: WO 0114535-A 1 01-MAR-2001;  
 Schwientek, Tilo (DK); Clausen, Henrik (DK)  
 Location/Qualifiers  
 1..1362  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..1362  
 /note="The gene of C2GnT3"  
 /protein\_id="CAC34674.1"  
 /translation="MTKFCYFKHTLOOKYVFLPTLMLSLKLVNRLFPKDI  
 LVEYSLSFVRNRYTHVDEVRVENCSEIYEOPLTEGSLERIRRIIDLEDD  
 VVAMTSDCDITGRTAKLVKSEESFPIASLVYHRAINVERIRIATHYQNHNY  
 CIHYDRKAPDTFKVAMNNLAKCFNSNIFASKLEAVEYAHISRLQDNLCLSLKSSI

Query Match 100.0%; Score 1362; DB 10; Length 1362;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 410 a 260 c 279 g 413 t

ORIGIN  
 QMKVYINLCGGDFPLKSNELVSELKLNANLETVKPPNSKLEERTYHEELRPVY  
 EYVKLIRINISKEAPPHNIQIFVGSATYFLVQAQFYIFNNSIVDFWMSKDTSP  
 DEHFWATLIRVPGIRPEISVADVDLOSIRFLVWNYEYEGFVMSKLRVSCI  
 YGAELRWLIRKDHMFANKEFDSVYDILIKLALEKLEDEQROWITLPSKRLPMENLIT  
 TS"

61 ttaacccatgagcgcctcctcttgtaaagcttctaagtgaagcagcctcttcgcaa  
 61 TTAAACCCATGAGCGCTCCTCTCTTGTAAAGCTTCTAAATGAGACAGCTTCCGCAA  
 121 aaagcattctcgttgatgagtaacctgaagccttcttgaaagaagatc  
 121 AAAGCATTCTCCTGATGAGTAACCTTAAAGCTTCTGTAAGAACAGATC  
 181 actcatgttaagatgaagcagatgaagtaagcttctgagatcctatgaacagag  
 181 ACTCATGTTAAGATGAAGCAGATGAAGTAAGCTTCTGGGATCTATGAACAGAG  
 241 ccttggaaattggaagaagcttgaagaatgaagaagcattcattgattgagat  
 241 CCTTGGAAATTGGAAGAAGCTTGAAGAATGAAGAAGCATTGATGAGGAT  
 301 gatgatgttgatgagcagcagcagcagcagcagcagcagcagcagcagcagc  
 301 GATGATGTGTGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
 361 caaagcctgtctcaagaagaggaagaagcctcccaatagccttcttgctgcac  
 361 CAAAGCCTGTCTCAAGAAGAGGAAGAAGCCTCCCAATAGCCTTCTTGCTGCAC  
 421 aaagatgacattatggtgaagaagccttccatgatacaaacacacacattac  
 421 AAAGATGACATTATGGTGAAGAAGCCTTCCCATGATATCAACACACACATTAC  
 481 tgcattcattatgacatgaagcagcagcagcagcagcagcagcagcagcagcagc  
 481 TGCATTCAATTATGACATGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
 541 aagtgcttcccaattatcttcttcccaattatgagcagcagcagcagcagcagc  
 541 AAGTGCTTCCCAATTATCTTCTTCCCAATTATGAGCAGCAGCAGCAGCAGCAGCAGC  
 601 tccagatccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc  
 601 TCAGATCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
 661 aaatagttataactgctgctgctgctgctgctgctgctgctgctgctgctgctg  
 661 AAATAGTTATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG  
 721 tcagattgaaanaactcagcagcagcagcagcagcagcagcagcagcagcagcagc  
 721 TCAGATTGAAANAACCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
 781 aaattggaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc  
 781 AAATTGGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
 841 ctcccaataaggaacaacatctccaagaagcagcagcagcagcagcagcagcagcagc  
 841 CTCCCAATAAGGAACAACATCTCCAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
 900

QY 901 ggcagtgctatttcttcaagcatttcaataatatcttcaaacactccatc 960  
 Db 901 ggcagtgctatttcttcaagcatttcaataatatcttcaaacactccatc 960  
 QY 961 gttcaagacttttctgctgtcctaagacacatactctcctatgagcacttttgact 1020  
 Db 961 gttcaagacttttctgctgtcctaagacacatactctcctatgagcacttttgact 1020  
 QY 1021 acccttgctgggttccagaataaccctgggagatctccagatcagcccaaggatgctc 1080  
 Db 1021 acccttgctgggttccagaataaccctgggagatctccagatcagcccaaggatgctc 1080  
 QY 1081 gatctgagagtaagactcgcctgttcaagtggagattaccatgaaggcttttctatccc 1140  
 Db 1081 gatctgagagtaagactcgcctgttcaagtggagattaccatgaaggcttttctatccc 1140  
 QY 1141 agttgactgagatcaccctcgaagcgtgtgtatttaaggagctgcagaataaaggctg 1200  
 Db 1141 agttgactgagatcaccctcgaagcgtgtgtatttaaggagctgcagaataaaggctg 1200  
 QY 1201 ctatcaagaatgagacattgttctgtaataatttgaattcctgaagggccctactctg 1260  
 Db 1201 ctatcaagaatgagacattgttctgtaataatttgaattcctgaagggccctactctg 1260  
 QY 1261 attaaatgcttggcagaagaacttgaagaacagcagagagacttgaacttgcctca 1320  
 Db 1261 attaaatgcttggcagaagaacttgaagaacagcagagagacttgaacttgcctca 1320  
 QY 1321 gaaaagtattattatgataagaatctcctaccatcatca 1362  
 Db 1321 gaaaagtattattatgataagaatctcctaccatcatca 1362

RESULT 2  
 AF132035 3435 bp mRNA PRI 08-APR-2000  
 LOCUS Homo sapiens core 2 beta-1,6-N-acetylglucosaminyltransferase 3  
 DEFINITION (C2GnT3) mRNA, complete cds.  
 ACCESSION AF132035  
 VERSION AF132035.1 GI:7527463  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 3435)  
 AUTHORS Schieler, T., Yeh, D.C., Levery, S.B., Keck, B., Merckx, G., van  
 Kessel, A.G., Fukuda, M. and Clausen, H.  
 TITLE Control of O-glycan branch formation. Molecular cloning and  
 JOURNAL characterization of a novel thymus-associated core 2 beta1,  
 Copenhagen, 6-N-acetylglucosaminyltransferase  
 J. Biol. Chem. 275 (15), 11106-11113 (2000)  
 MEDLINE 20219156  
 REFERENCE 2 (bases 1 to 3435)  
 AUTHORS Schieler, T. and Clausen, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-FEB-1999) School of Dentistry, University of  
 Copenhagen, Nørre Alle 20, Copenhagen 2200, Denmark

FEATURES  
 source  
 1..3435  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..3435  
 /gene="C2GnT3"  
 862..2223  
 /gene="C2GnT3"  
 /codon\_start=1  
 /product="core 2 beta-1,6-N-acetylglucosaminyltransferase  
 3"  
 /protein\_id="AF63156.1"  
 /db\_xref="GI:7527464"  
 /translation="MKIFKCYFHKTLQCKVFLFLTLMLSLIKLNNRLEPQKDIY  
 LVEYSLSTSPFVANNRTYHWKDEVRYEYVNSGIYEDQPLEIGKSLRERDIDLEDD

BASE COUNT 1032 a 678 c 651 g 1074 t  
 ORIGIN  
 VVAMTSDCDIYQTLRGYAOKLVSKSEKSFPIAYSLVYHKDAIWERLHAIYQNHIT  
 CIHYDRKAPDPEFKVAMNNLAKCFENIFLASKLEAVEAHISLRQADINCLSDLSKI  
 QMKVYINCGDDPDKSNFELVSEKTKNGNMELVYKPPSKLERFVYHLEIRVY  
 EVKLTPTNTSKRAPHNIOIFVNSAVYLSQAFVYIFENS.VVOEFMWSKQYSP  
 DEHRATILIRPPIGELSRSAQVSDLOSTRLVKNMYKGFEPYSGSHLSRVCI  
 YGAELRWLIDGHWFAKFDPSKVDPIIKCLAKLEBQCDWITLPSKLFMRNLT  
 TTS"  
 Query Match 100.0%; Score 1362; DB 89; Length 3435;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atgaagatctcaaatgttattttaaactacccctacagcagaagatttccactgttt 60  
 Db 862 atgaagatctcaaatgttattttaaactacccctacagcagaagatttccactgttt 921  
 QY 61 ttaaccctatggtcgtcctcttgtttaaagcttcaaatgtgagagactcttcgcaa 120  
 Db 922 ttaaccctatggtcgtcctcttgtttaaagcttcaaatgtgagagactcttcgcaa 981  
 QY 121 aagaacatttacttgggttgaagctacccctaaagactcgttttgaagaacagatc 180  
 Db 982 aagaacatttacttgggttgaagctacccctaaagactcgttttgaagaacagatc 1041  
 QY 181 actcatgttgaagatgaagtcaggtatgaagttactgttcgggtctcatgaacaggag 240  
 Db 1042 actcatgttgaagatgaagtcaggtatgaagttactgttcgggtctcatgaacaggag 1101  
 QY 241 cctttgaaatttgaagagctcgtgaagaataagaagaaggagacatctgaactgagagat 300  
 Db 1102 cctttgaaatttgaagagctcgtgaagaataagaagaaggagacatctgaactgagagat 1161  
 QY 301 gatgatgttggcgaatgacagtgatgtgacattatcagaactcagaaggttactgct 360  
 Db 1162 gatgatgttggcgaatgacagtgatgtgacattatcagaactcagaaggttactgct 1221  
 QY 361 caaagcttgcctaaaggagaggaagaagcttcccaatagcctattcttgggtgtccac 420  
 Db 1222 caaagcttgcctaaaggagaggaagaagcttcccaatagcctattcttgggtgtccac 1281  
 QY 421 aagaatgcaattatggttgaagaagcttaccatgctatatacaacagcacaatatttac 480  
 Db 1282 aagaatgcaattatggttgaagaagcttaccatgctatatacaacagcacaatatttac 1341  
 QY 481 tgcattcatttgcctgaagcagcctgataaccttcaagattggcagacaatttgaact 540  
 Db 1342 tgcattcatttgcctgaagcagcctgataaccttcaagattggcagacaatttgaact 1401  
 QY 541 aagtgcctctccaatatttccattgtcttccaattagaagcgtgtgaaatagccacatt 600  
 Db 1402 aagtgcctctccaatatttccattgtcttccaattagaagcgtgtgaaatagccacatt 1461  
 QY 601 tccagactcagagctgatttaattgtcttgcggaacttccgaaggttccaatccagtg 660  
 Db 1462 tccagactcagagctgatttaattgtcttgcggaacttccgaaggttccaatccagtg 1521  
 QY 661 aaatagtatcaacttctggtggaagaatttccctcgaagtcgaatatttgaattgtg 720  
 Db 1522 aaatagtatcaacttctggtggaagaatttccctcgaagtcgaatatttgaattgtg 1581  
 QY 721 tcaagagtgaaaaaactcaatgaggaataatgtttggagagcgtggaaccccaaaagt 780  
 Db 1582 tcaagagtgaaaaaactcaatgaggaataatgtttggagagcgtggaaccccaaaagt 1641  
 QY 781 aaatggaagaatcactacatcatgaacttgaagcgtgacttataatgattgtgaag 840  
 Db 1642 aaatggaagaatcactacatcatgaacttgaagcgtgacttataatgattgtgaag 1701  
 QY 841 ctacaataaggacaacactcgaaggaagcaccctccataaactcgaatatgttct 900

Db 1702 CTACCAATTAAGGACAAATCTCCAGAGAACACCCCCATTAACATTCAGATATTGTT 1761

Qy 901 ggagctctatttctgtttaagcaagcattgtttaataattttcaacaactccatc 960

Db 1762 GGCAGTCTATTGTTGTTTAAGTCAAGCATTTGTTAAATATTATTTTCAACAACCTCATC 1821

Qy 961 gtcaagaactttttgctgtctcaagaacacatactctctcatgagcaactttggact 1020

Db 1822 GTTCAGACCTTTTGGCTGGCTTAAGACACATACCTCTGATGAGCACTTTGGGCT 1881

Qy 1021 acctgattcgggttccaaggaatacctggggagatttccagatcagcccgagtgctc 1080

Db 1882 ACCTTGATTGGGTTCCAGGAATACCTGGGAGATTTCAGATCAGCCCAAGATGTCTC 1941

Qy 1081 gactcgaagtaagactcgcctgtccaagtgtgaattactatgaaggcttttctatccc 1140

Db 1942 GATCTGCAGATAGACACCGCTTGTCAAGTGAATTAATGAGAGGCTTTTCTATCCC 2001

Qy 1141 agttgactgattcaccctcgaagcggtgtgtattatgagagctgagaattaaggttg 1200

Db 2002 AGTTGTACTGATCTCACCTTCGAAAGCGGTGTATTATGAGCTGCAGATTAAAGGTG 2061

Qy 1201 cttaacaagaatggaactgtgttgcataaattgattcttaaggtgagaccatcttg 1260

Db 2062 CTATCAAGATGACATGATGTTGTTGCTAATTAATTGATTTCTAGGTGAGACCTATCTTG 2121

Qy 1261 attaatgtcttgagaagaagcttgaaagacagagagaagactgagatcactttgacctca 1320

Db 2122 ATTAAATGCTTGCGAGAAAGCTTGAGAACAGACGAGAGCTGATCACTTTGCCCTCA 2181

Qy 1321 gaaaagtattatgatgataaattctactaccacacata 1362

Db 2182 GAAAGTTATTATGATGATGAATCTCCTACCTACCATCATCA 2223

RESULT 3  
AB037596 3508 bp mRNA ROD 20-OCT-2000  
LOCUS Mus musculus Igt B mRNA for  
DEFINITION beta-1,6-N-acetylglucosaminyltransferase B, complete cds.  
ACCESSION AB037596  
VERSION AB037596.1 GI:9650953  
KEYWORDS beta-1,6-N-acetylglucosaminyltransferase B,  
SOURCE Mus musculus (strain:ICR) 13 day embryo fetal brain  
tissue.lib:lambda ZAP cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Chen, G.Y., Kurosawa, N. and Muramatsu, T.  
TITLE A novel variant form of murine beta-1,  
6-N-acetylglucosaminyltransferase forming branches in  
glycobiology 10 (10), 1001-1011 (2000).  
JOURNAL 20485361  
MEDLINE 2 (bases 1 to 3508)  
REFERENCE Kurosawa, N., Chin, G. and Muramatsu, T.  
AUTHORS Direct Submission  
TITLE Submitted (25-JAN-2000) to the DDBJ/EMBL/GenBank databases.  
JOURNAL Nobuyuki Kurosawa, Nagoya University School of Medicine, Department  
of Biochemistry, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550,  
Japan (E-mail: nkuros@suru.med.nagoya-u.ac.jp, Tel: 81-52-744-2063,  
Fax: 81-52-744-2065)

FEATURES  
Location/Qualifiers  
1..3508

Source  
/organism="Mus musculus"  
/strain="ICR"  
/db\_xref="taxon:10090"  
/dev\_stage="13 day embryo"  
/tissue.lib="lambda ZAP"  
/tissue\_type="fetal brain"  
1009..2214  
/gene="Igt B"

CDS 1009..2214  
/gene="Igt B"  
/codon\_start=1  
/product="beta-1,6-N-acetylglucosaminyltransferase B"  
/protein\_id="BAB03495.1"  
/db\_xref="GI:9650954"  
/translation="MGSWKYSLEFSLIALMLMFWTDRIKIMKHYHPPRAVNSIVLA  
EVCLOMFGESEFYADSAKRTTLENFTCPKXIONNHITTELTSEEARFPLATLTH  
KDYDFERLFPRAIYMPONVYCVHVDSDATDFKFAVQLSCFPNLAASMEPVYG  
GESRLQADLNCMDLVASKIPMKVYLNTCGDPEPLKTKELVYKRFIKNLTPVYL  
PPAAVGRKTYVHQLDHRKPYHNARLKAAPPNNITTYFGTAVALRREFANVYL  
KDOESVPLISMSKDPYSPDEFHWYNTLRIPGVPSMNPMSMGNLAAVMMDEAKH  
GCGSHYHGICITGNDLQWLINSOSLFRANKFELNTPYPLTVELELRLENTLNQSE  
IALDPSWTF"

BASE COUNT 877 a 872 c 867 g 892 t  
ORIGIN

Query Match 15.2% Score 207.2; DB 94; Length 3508;  
Best Local Similarity 52.9%; Pred. No. 1.3e-39;  
Matches 505; Conservative 0; Mismatches 428; Indels 21; Gaps 2;

Qy 337 taticgacctlaagaggtatgtctcaaaagctgtctcaaaagagagaaagcttccca 396

Db 1231 TACAAGATTCAAAACACATACATTAACAGACTCTCTGAAGAGAACCCGCTTCCCT 1290

Qy 397 atagctattcttgggtgtgtccaaagatgcaattatgtgtgaaaggttatcatgct 456

Db 1291 CTGGCTTACACTTACTACTTCCAAAGATATGACACTTTGAGAGACTCTTCAGGGCC 1350

Qy 457 atatacaacagacacataatttactgcatcacaattatgactgtaagagcactataccctc 516

Db 1351 ATTATCATGCCCCAGACAGCTACTGTGTGACAGTGAGATAGTAAGCGCACACACTTTC 1410

Qy 517 aaagtgtcatgacaattatgactgctctctccaaatttctattgtctccaatta 576

Db 1411 AAAGAAACGGGTGGCGGAGTTACTTAAGCTGTTCCCAATGCTCTGCGCTCTTGAGATG 1470

Qy 577 gaagcgtgtgaatgccccacatttccagaccagcgtgacttaattgtctgtcgagac 636

Db 1471 GACCCGGTGTCTATGTGTGGCTTCTCCGGCTCCAGGCTGATCTGAATGATGAAGAT 1530

Qy 637 ctctctgaagcttcaatccaagtggaataatgttatcaacttctgtggggaagatttccc 696

Db 1531 CTGTGTGGCTCCAAAGATCCCTGGGAATATGTCTTCAACACCTGGCGGAGGACTTTCT 1590

Qy 697 ctgaagtcacaatttgaattgtgtcagaggttgaaaataactaatgagacaataatgtgtg 756

Db 1591 CTGAATAACCAAGAAAGAAATAGTTACGTACCTGAACGGTTATTGGGAAGAACCTCAC 1650

Qy 757 gagaaggttgaaccccccaacagtaaatgtgaaagttacttaccatcatgaacttga 816

Db 1651 CCAAGGGTGTGTGCTCTCCACATGACATGAGTGAAGAACCAAGAGCTCCACACGAGCTG 1710

Qy 817 cgggtgcttatgaatatgtgaagctacacaataagagacaacatcccaagaagacccc 876

Db 1711 TTAGACCATTAATAATCCCTACGTGC-----ACAATAACAGCAAGATTAAAGCTCCCA 1764

Qy 877 cccataacatcaagatatgttggcagtgcttatttggtttaagtaagcatgtgt 936

Db 1765 CCTTCACACCTGACCATTTACTTGGACCTGATTAGTGAGCTCTCACAGGGAATTTGCT 1824

Qy 937 aaataatatccaacaactccatcgttcaagacttttggcctgtgctaaagacacatac 996

Db 1825 AACTTTCTCTCAAAAGACCGCTTCAGTAGATTAATCCCTGGTCCAGGACAGAT 1884

Qy 997 tctctgaatgaagcatttgggttaccctgtgattcgggttccaggaataacctgggagatt 1056

Db 1885 AGTCTGATGAGCATCTTGCGGTGACACCTCAATAGATCCCTGGAATCCCTGCTCCAG 1944

Qy 1057 tccaagatcagcccaagatgtgtgtgactcgcagagtaagactgctgtcaagtgtgaat 1116

Db 1945 CCACCAACGCAATCCTTGAGAGGGGTAACTCAGAGCTGTGAGTGAATGACATGGAAGCG 2004



```
OY 1117 tactatgaagctttctctatccagcttgatcctcacccttcgaagcgtgtratt 1176
Db 2005 AAGCATGGAGGC-----TGCATGGTACTACGTCCTGATGCAATTGATC 2049
OY 1177 tatgaagctgcgaataaagctgcttaccgaagatggacattggttgcataaatt 1236
Db 2050 TATGGAACGAGACTTCCACTGGTGATTAATTCGAAACCGCTTGTGCAACAAATT 2109
OY 1237 gattcctaagtgcacctatcttatttaaatgcttgcgacagaagcttgaaga 1290
Db 2110 GAACACACATACCTCTTACCGTGAATGCTGGAATGAGGCTTCGAGAA 2163

RESULT 4
AF318573 AF318573 108873 bp DNA VRL 14-FEB-2001
LOCUS Bovine herpesvirus 4 long unique region, complete sequence.
DEFINITION AF318573
ACCESSION AF318573
VERSION AF318573.1 GI:12802528
KEYWORDS
SOURCE Bovine herpesvirus 4.
ORGANISM Bovine herpesvirus 4.
REFERENCE 1 (bases 1 to 108873)
AUTHORS Zimmerman,W., Broll,H., Ehlers,B., Bunk,H.-J., Rosenthal,A. and
Goltz,M.
TITLE Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
JOURNAL Identification of an Origin of DNA Replication
PUBMED J. Virol. 75 (3), 1186-1194 (2001)
REFERENCE 2 (bases 1 to 108873)
AUTHORS Zimmerman,W., Broll,H., Ehlers,B., Bunk,H.-J., Rosenthal,A. and
Goltz,M.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Project p24, Robert Koch-Institut, Nordufer
20, Berlin D-10555, Germany
FEATURES
source 1..108873
location/Qualifiers
/organism="Bovine herpesvirus 4"
/db_xref="taxon:10385"
/complement(445..633)
/note="long unique region; LUR"
CDS misc_feature
complement(445..633)
/note="ORF 1"
/codon_start=1
/evidence="not_experimental"
/product="hypothetical protein pbo1"
/protein_id="AAK07921.1"
/db_xref="GI:12802529"
/translation="MEGDVSVAFVRKRTRODLSNAGSNEYCAVLRKHNVSAGC
CAVKNRSHSLPGMPP"
614..4480
/note="tegument protein; ORF3 BORF1"
/codon_start=1
/product="v-FGM-synthase"
/protein_id="AAK07922.1"
/db_xref="GI:12802530"
/translation="MDTSPSIOLCFVSRAPLNSNCOVIOOFTSDGOLPLTVSLQY
BOYFVYLVNGEDAVRFVPIPLDSESSGELTPRESSILKTDSPRLSFTIGVSPA
PETALSRDLVALIQLSMGSLAPNSNLITIKIYSLISIKNLNPSRKHMSTLRAL
CPDPSLFRYKGESEIVGAFTHGAPQGLGSEPKGLDFAFGGLNAGSGSNGMT
POHQPOTIVNONGKATYIVPDYWIPELSTYLSGHTDGLAVTGOALVARDN
LKFLYNSNMGKNGVYSSQGLASTVSTLSSRSSELPYTDIMDAVINNRCK
SLPCTQGYEPPIVESHLPDLSVEEPIATGAVCESRNITDLPETRKINIVLGT
FTSHSDSREVCILGALKYLSFLRPSQVLSAVKCEHMAKLLSVCGTIGIK
TYSALPRLNGLPETAASNRQNTMTKOHNLIVAVIATADPNTDITVLK
AGCAIHGCOIKVLGRTOQPHITIVDKKQENVRKISIKRSPAPLSITKLKOS
ENWQDSSMDWDLPSLSTLELILHFAVGCKDFIVKHIDRLSSGVNAQCGIGARDI
PIDVGIIVSDLSLAAPDGSSENPWATASLDDPLVNNEPVPGIGSLAGECPPLS
TYPPIGAHVAIVESLNLIVSAFPRKRIDVTCTFWPQITDSHGLILKRAEEL
CSQIGICVFTSCTSNRRGQKCLNLSLVRYLIVATAPCKDVTQGLTPDLKEPDSI
```

gene

CDS

CDS

CDS

CDS

```
VWLPSTIEYHTFGVVISOLFRDPSGCHINIDPLVYKLLITVTKHINHEISICHNV
GCGGLITACFEEMAYAGGASIALITVPQDDPVLITSETPLGAVEPVRYKSTYOKHLE
TSDIIEPVNGRTLPESVANSTFTVSKRELIIFRESINENVENWNRKFSREQRYPCRY
OQTPEDPDKDLHLTLPQSMCPHGPYKHQOVNYYLLPGVINTPDLVLAEEAEFRVA
LSTPTDKTKIVITDTTNGVGLIGANIDALIGDAIKAMYTRKNSVLAIFEPENH
NSDPYSLAIGTACOLLENNKFMGNKNSNTTMCCKENSGLMESRMLNFIPEHTH
AVALDSMGSLPLPGTIOGTHLGFAPDSITVEMSLTHGMVATOFQAGISAPALYTP
QNPRTGYITISGCSADGRHLALHDPLGSLNNLMQPHIIPKATPLAKSPKRMRLDLH
IWANKYREMDPPPHDPDLRIKVM"
4608..4811
/note="ORF 2"
/codon_start=1
/evidence="not_experimental"
/product="hypothetical protein pbo2"
/protein_id="AAK07923.1"
/db_xref="GI:12802531"
/translation="MASPTQYCHRWVRYHGPRKSAKISYGIKRDYTHQLTSKNSAYS
GMSQPCSCADLNKLFVSQCPIS"
5286..5513
/note="ORF 3"
/codon_start=1
/evidence="not_experimental"
/product="hypothetical protein pbo3"
/protein_id="AAK07924.1"
/db_xref="GI:12802532"
/translation="MGFNPMVLYFRPCNCTPRLKHGILYIKYISYLIRATFLVRVY
HGSPKRDTRPPGPGPPELSITVQGLAKK"
5943..9347
/note="ORF 4"
/codon_start=1
/product="single-stranded DNA-binding protein MDPB"
/protein_id="AAK07925.1"
/db_xref="GI:12802533"
/translation="MAMTKAAGQGYEDNIGSKAPIGPCGYIYKKNPPTREASLIG
NCPGGSVNSVPLRGLIYVPPGPRIVKVKHIDDTTAVKTSYHREVIENSNDC
FPTFHTGICPOLCOTRDLFGTTHSPNSNTNITSLDGLGTDGIDEMILGVIYT
EFGFERLYHSHMPLIGHVEVVISQAHARKPIPLYDEDLFSIARDMKRRHSVDS
QYLENSLYTAIAQIIVKDAVATIQMESQFVNDQMKRVYKRRKPSSSSGSDGL
SLIMDSVASELAVSYGLFIDAPQMSALDITSPITVDCTEDRQALANRKLK
OSIHNTOLESTSMILYTRIQONKRPATKGLDYNVSYLQHGSLYSEAVQDEGQ
PVFOGQSNLSGSTYTIHLAAVAMCPPTLARKCYLQFCQHSQTNQSNINISY
VGRASDMQNLGCKRPACINTFLPRDRPPIITTSRRDPYITIGSGWYNIDLF
ILGNFASFRKEDBGQAVQEVCKYTWOLITLLELKGITDTPPEKLVADISF
IKVFKDIDAVDIEVLFINKCKNNINRETIKSHHYVQYCNCFAPPCGVFOL
YKSLITIIDVCLPNCMEDEDNSSMGMSPLMKNHOTOITNKKCCGCPDGVYTG
AELKVMHSDPCDFDVAIDINGISPTKQVARSAMLVERTIKIRNITFSNBSG
SEALOSGLKSGSKKNYIVTGYMKFLINYHRIKMPOTKMSLFMWHFSTKQIPL
ITGVKQDLVALANYIENSKLSEIDVIDIIPDNITVAKIKLNNAIILACQTOFY
STPHICLIPVRYVPGREYHVLNEEFISTEYQIKVQNRQIQAQILKEDVQAQMG
KLRIITVPMVINKYNGINGNIGIFHCGNLGYMGAGVDENLIFENAPVROSTNAM
KRRHVENTPYDNLIKRTYTPSTPSTPEVNIIRDIMLEDKRONNIFRDVYELVKG
LQCAQADLADDDIOFYIGERYINSDLSLOSISDNGVWCQSDVSPFLGFTVSEB
QLEPIGEQOTSPASIDEDFFQASGLSTIAAGKRRKLNCLSPDL"
9352..11430
/note="ORF 7"
/codon_start=1
/product="transport protein"
/protein_id="AAK07926.1"
/db_xref="GI:12802534"
/translation="WABDLAVYSQVDFLSKTSLSLTFCDPRHIDGAPORNRKRLK
LIDLALYSLVQNRTECSCLSELEHLINITYITQWLETCNLNTEKRYFPMHLHS
DLCCFRTKYNLCFEGDCVIPSLSIINDIEFRLNVSVCYSRNAALGEBETINL
LGTLRNISPILPLDLYPSISCMNCLNETSLIPNAGNVTIINMSIKTEKVAIPVABE
PVQGLFENEHLHLINIPDQGNKQULESTLIPNAGNIEAIDQNIIEKASQIILEIS
NLITYNSHGSIDBPKISCSQMATLIIHETEMASHRQKVASHLATLPPIHFDPSYA
YPIESLFCGSIIFYIIMNDIQAIDKDCSTETPAKSNYETIKERKEIVLRNLNLABS
TREGBIDEPNYSIQARHTIGETSEQISDAAKRNNYLQKATTEGCLKIDTDEET
QKIKMNCNLTLRTGMVYIKAAILKHNFMIRKFFILPQMHHTYINELYESKTI
KNSLYLQISOEHIDISITQYSLITGLVGPFPFSPPIILLANCLDARAKPHK
MLISEVIMPMSPEKMDIABENAFNYINTGKLNVOOCAMRYIELVSVSLNVRWE
KQINIVSSTLSLPHNLSPKPTFRKSGIYIYEKTAAPLFIHNDTGMIFKDLVALLVH
LQLSHKRYVL"
11420..11404
```

CDS	<p>/gene="gb" 11420..114044 /gene="gb" /note="ORF 8" /codon_start=1 /product="glycoprotein B" /protein_id="AAK07927.1" /db_xref="GI:12802535" /translation="MYKTLFPALIKVCSFNQTTHTSTTSPSISSTTSSTTSK PSNTSTNSLASPONTSTSKPSIDNOSTSPITPTVVDIASNFKRYCSASS SGELFFDLQCPDPTDKKHVEGILLVKNIVYIFKRYKRIATSVYRMSO AAVNRDDISRAIPNEISIDRTYCFSAFATVINGIINTYIDDSERKSVLPVA GLTENNRFSOPLIYAEPGPGIYRVATVNCVEVDVAVSVPYTHFALDIT ELSPCHNNSOCTHGTSTSRDAKVIENHQTVDYERGHPTKDKRIEIKDEETIS WKREDERAIQDPVTKTTPRAIQTHTNSFPFVANEVYASLTISNOSETRGTEI LNCNSTINETLEEVYKFNKSHIDGKVKYIKINGLEFLIQAKPLSEHTVYTI ERNKTKGNSRKRQSVDTKFOGAKGLSTAOVYAVDHLRTSMNHILELITWCREQ KKNLMWYELSKINPVSAAYGKPVAKAGDAFEMSECINVDASVNIKSMRTD DKPVYSRLVTFKFNSTATERGLTNEILLNTHECTRPTADHYFEVKNATHY FKRYKFKTMDNTNISTDPLTLNLTEDIDTFKVELYSETERMSALDETFR EYKYVOKIASREDIDNTIDNRDVIKDISEMADLGDIGKVVVNTESGIVTYES IVGGEVSPFTNIGVTTIILLIVYFVYISRRNNNEAPIKIYVINDKASBOE NIQPLPGEETKRILLDHODQSEHGKSEASHKGLQLLDGLDLRRNGYTRLP TFDPSGNDTSETHOKYV"</p>	<p>Matches 520; Conservative 0; Mismatches 427; Indels 39; Gaps 3; QY 316 atgaccagtgattgtgacattatcagactcctaaggtatgctcaaaagctgtc 375 Db 107637 ATAACGAGAGCTGTGACGCGCTTCAAGGCCCAAGAGATTCAATGATTCACACTAGT 107696 QY 376 aaggaggaagaagctcccaatagacctatctcttggtgtgcacaagaagtgaactatg 435 Db 107697 AAGAGAGAGCTAGACTTCCCATTCCTACCTGATGATGGATGAGAAATTTGAAAC 107756 QY 436 gttaagaagctatccatgctatatacaaacagacaatatctcgtcatccattatg 495 Db 107757 TTGGAACGCGCTGCTCGAGCGGTGTATCCCTCAGAACATATCTGTGTCCAGTGAT 107816 QY 496 cgaaggagacctgatacctcaaatgtccatgacaacattagctaagtgtctccaat 555 Db 107817 GTGAAGTCCCGAGACACTTTCAAGAGCGGTCAGAGGCATTATTCTGCTCCCAAT 107876 QY 556 atttcattgcttccaatagaggctgtggaatatgccacattccaagctcaggct 615 Db 107877 GCTTCATGCGCAGTAAGTGTGTCGGTGTATTCCTGCTCCAGAGTCAAGT 107936 QY 616 gatttaaatgtctgtcgagaccttcgaagcttcgaatccagtggaatatgtatac 675 Db 107937 GACCTGAAGTGTATGAGAACACTTGTCTCCAGAGCTCAGTGCATGAACTACTTACTGAT 107996 QY 676 ttgtgtgggcaagatttcccttgaaagtcacatttgattgtgtgcagagttgaaaaa 735 Db 107997 ACATGGCGGACAGACTTCCCATTAAGAACCAATGCCGAGATGTCCTGCTCCAGATG 108056 QY 736 ctcaatggagcaaatatglttgagagagcgtgtaaaccccaagaagaattgaaagctc 795 Db 108057 TTGAAGGTAAGAACGTAATGAGAGTGTGAGTACCTTCTGAGTCCAAAAAATGCTGG 108116 QY 796 acttacacatgaacttagagcgggtgccttataatgtgaatgtgaagctacaataagaca 855 Db 108117 AATATCCGCTATGAGGTGACAGACACACTGTACCTT-----ACG 108155 QY 856 aacatctcaagaagaagcaccaccccaatacctagataattgttgagcagtgctattt 915 Db 108156 AGCAAGATGAGAGACCTCCCTCGTAATTTACCATGTTACAGAGGATCCATTTT 108215 QY 916 gtttaagtcgaagcattgttataatatattccaacaactccatcgltcaagaactttt 975 Db 108216 GTGGCCTGTGACCTTTTGCAACATGCTTAGAACACCTAAATCCCAAGACTGGTT 108275 QY 976 gcttgatctaaagaacatactctctgtatgagcaatttgggtacacttgatcgggtt 1035 Db 108276 GAATGGGTCAAAAGACACTATAGCCCGGAGAACACCTCTGGGACACCTTCAAGCTGCT 108335 QY 1036 ccaagaaatacctggagagatttccaatcagcccaagatg---tgtctatctcagaagt 1092 Db 108336 CCGGATGCTGCTGCTGTCTGTCTAGCACACCCCAAGTATCACATGCAAGACTGCGC 108395 QY 1093 aagaatcgctgtcgaagtgaattactatgaagctttt-----tat 1137 Db 108396 ATGGCCAGGCTGCTGCTCAAGTGGCATCACAGAGGAGATGTCATGGGGGCGCTTAT 108455 QY 1138 cccagtgtaactgatactcaccctcgaagcgtgtgattatgagctcaagatlaaag 1197 Db 108456 GCACCGCTGCTGGAATCAATCGGAGGCGCATGTGATTAGCGGGCCGGGACCTGTAC 108515 QY 1198 tggcttaacaagaatggaactgttctgataataatttgatctcaagtgtgacctatc 1257 Db 108516 TGGATTCTCCAGAACATCACTCTTGGCAAAAGATTGACCGGAGGATGATTAAC 108575 QY 1258 ttgattaaatgctggcagaagaagct 1283 Db 108576 GTCTCGAGTGTCTAGAACAGTACT 108601</p>
-----	---	--

Query Match 14.3%: Score 194.8; DB 58; Length 108873;  
Best Local Similarity 52.7%: Pred. No. 1.6e-36;

LOCUS AB037597 1206 bp mRNA ROD 20-OCT-2000  
 DEFINITION Mus musculus IGF1 A mRNA for beta-1,6-N-acetylglucosaminyltransferase A, complete cds.  
 ACCESSION AB037597  
 VERSION AB037597.1 GI:9650955  
 KEYWORDS beta-1,6-N-acetylglucosaminyltransferase A,  
 SOURCE Mus musculus cell\_line:F9 cell CDNA to mRNA.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (sites)  
 Chen, G.Y., Kurosawa, N. and Muramatsu, T.  
 A novel variant form of murine beta-1,6-N-acetylglucosaminyltransferase forming branches in poly-N-acetylglucosamines  
 Glycobiology 10 (10), 1001-1011 (2000)  
 JOURNAL 20485361  
 MEDLINE 2 (bases 1 to 1206)  
 REFERENCE Kurosawa, N., Chin, G. and Muramatsu, T.  
 AUTHORS Direct Submission  
 TITLE Submitted (25-JAN-2000) to the DDBJ/EMBL/Genbank databases.  
 JOURNAL Nobuyuki Kurosawa, Nagoya University School of Medicine, Department of Biochemistry, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan (E-mail: nkuro@tsuru.med.nagoya-u.ac.jp, Tel:81-52-744-2063, Fax:81-52-744-2065)  
 FEATURES  
 source  
 1..1206  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_line="F9 cell"  
 1..1206  
 /gene="Igf1 A"  
 /gene="Igf1 A"  
 1..1206  
 /codon\_start=1  
 /product="beta-1,6-N-acetylglucosaminyltransferase A"  
 /protein\_id="BAB03496.1"  
 /db\_xref="GI:9650955"  
 /translation="MPSRVRFVFSVTVTVFIVLVSPGDSQYOKLINSVRL  
 SOVCSFIDGSRFLRMKIMHRSKSTVEVTSHTTAPLSEVDEPLAIVYVH  
 HNFDIFARLFRATPONTICVHDEKTAFFKAVEDLYVCFPNALASMPYVVG  
 GISRLQADLNCIKIDLTSEVPEWKAINTGQDFPKTKKEIVYLKIGKNTLPGL  
 PPAHIGRTYRVHREHLSKELSYVIRITALKPPPHNLTIVGSAVVALSEFANFVL  
 RDPRAVDLHMSKDTFSPDEHFWYTLNIPQVSPMPNASTGTNLRAVKMDWEASH  
 GGCHGVHGIYICVINGDQLWLINSQSLFANKFELNTPLVIELELRERTLNQSE  
 IAIQPSWYF"  
 BASE COUNT 300 a 318 c 285 g 303 t  
 ORIGIN  
 Query Match 14.3%; Score 194.4; DB 94; Length 1206;  
 Best Local Similarity 52.1%; Pred. No. 1.5e-36;  
 Matches 497; Conservative 0; Mismatches 436; Indels 21; Gaps 2;

QY 637 ctctgaagcttcaatccagatggaaatctgtatcaactctgtggcgaagatttccc 696  
 DB 523 ctctgaagcttcaatccagatggaaatctgtatcaactctgtggcgaagatttccc 582  
 QY 697 ctgaatcaatatttaattgtgtcagagtgtaaaaactcaatggcgaatattgt 756  
 DB 583 ctgaatcaatatttaattgtgtcagagtgtaaaaactcaatggcgaatattgt 642  
 QY 757 gaaacggtgaaccccaacagtaaatggaaagattcaactacatgaactaga 816  
 DB 643 cggggggtgtgtctcagacccacgacattggaagacagagatc-----gtccacccg 696  
 QY 817 cgggtgtcttaataatgtgaagtcaccataaagacaaactcccaagaagacc 876  
 DB 697 gaaacggtgaaccccaacagtaaatggaaagattcaactacatgaactaga 756  
 QY 877 cccataacatcaagatattgttggaagtgctatttggtttaagtaagcattgt 936  
 DB 757 cccataacatcaagatattgttggaagtgctatttggtttaagtaagcattgt 816  
 QY 937 aaataatttcaaacatccatgcgtcaagacttttgcctggtctaaagacatac 996  
 DB 817 aaattgttctccgtgacccacggcggtgatttgcctcattgttccaaagattttc 876  
 QY 997 tctctgaagcactttgtggtacattgattcgggttccagaaatcctggggaatt 1056  
 DB 877 agtccgattgagcatttctggtgagcctcattgagattccagagattcctgctcag 936  
 QY 1057 tccagatcagcccaagatgtgtgtatctcgaagtaagactgcttgcgaagtgaat 1116  
 DB 937 ccac 996  
 QY 1117 taactaagagcttttctatccagttgactgactgactacattcgaagcgtgtat 1176  
 DB 997 aagcattgagc-----tgccatggttgcacacacacacacacacacacacacac 1041  
 QY 1177 tatgaagctgaagaaatgaagtgcttcaataagaatgagcattggttgaataatt 1236  
 DB 1042 tatgaagctgaagaaatgaagtgcttcaataagaatgagcattggttgaataatt 1101  
 QY 1237 gatcctaagtggaagccttcttgaataatgcttgcagaaagcttgaagaa 1290  
 DB 1102 gaactcaac 1155  
 RESULT 6  
 AX087948 1317 bp DNA PAT 17-MAR-2001  
 LOCUS AX087948  
 DEFINITION Sequence 14 from Patent WO0114535.  
 ACCESSION AX087948  
 VERSION AX087948.1 GI:13396926  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 1317)  
 Schwientek, T. and Clausen, H.  
 UDP-N-acetylglucosamine: galactose -4(1)-3-N-acetylglucosamine -9(a)-1-(/galactose to galnac) -9(b)-1,6-N-acetylglucosaminyltransferase e, c2gnt3  
 Patent: WO 0114535-A 14 01-MAR-2001;  
 Schwiientek, Tilo (DK); Clausen, Henrik (DK)  
 JOURNAL Location/Qualifiers  
 1..1317  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..1317  
 /note="The gene of C2GNT2"  
 /codon\_start=1  
 /protein\_id="CAC34676.1"  
 /db\_xref="GI:13396927"  
 CDS  
 QY 577 gaggctgtgaaatagccacattccagactccagcgtatattgaattgcttgcggac 636  
 DB 463 gaggctgtgaaatagccacattccagactccagcgtatattgaattgcttgcggac 522

```

/translation="MYOMKRIQOLHYLWAGCYMLATVALKLSPELKCDSHDLGLES
RESQSYCRNIIYNFLKLPARKSINSQVTRGQOEVLQAIINLNNEVKKRPPEDTH
YLSLTDCEHFEKAEKFIOPLSKEVEEPFLASVYHKEIEFELRLAYAPONIT
CYAVDEKSPETKEKAEKFIOPLSKEVEEPFLASVYHKEIEFELRLAYAPONIT
PKMYPLNTGCTDEPKISNEMOALKMLNGRNSMESEVPEKRETRMYHFEVBDTL
HLTKRKDDPPYPLNLTFTGNAYIVASRDEVOHYLKNPKSQOILEMYKDYSPDEHLM
TLORARMPGSVPHNPKYDIDSMTISARLYVKOGHGBDIDKGPVAPCSGISHORALICV
YGAIDLNMWLNQNHLLANKRDPKVDNNAQCELYRYAIVGTTEL"

```

```

BASE COUNT      359 a      307 c      322 g      329 t
ORIGIN

Query Match      14.1%; Score 191.8; DB 10; Length 1317;
Best Local Similarity 52.5%; Pred. No. 6.4e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

Qy 317 tgaccagtgattgacattatagaagctcaagaagttatgctcaaaagctgtctcaa 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 TCACCAGAGCTGTGAGCATTCAAGGCTGAAGAGAGTTCATACAGTCCCACTGAGCA 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 377 aggaaggaagaaagcttcccaatagcctattcttggttgtccacaagaatgcaatla 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 AAGAAGAGCTGAGAGTTCCTTATTCATCTCTATGCTGATCATGAGAAGATTGAAC 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 437 ttgaaggctatccatgctatatacaacgaacaaattactgcatccatataatgctc 496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 437 TTGAAGAGCTACTGCGACCTGTGTGCGCCCTCGAAGCATATGCTGTCATGTGATG 496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 497 gtaaggaacccgatataccttcaagttgccaagaacatttggcaagtgtctccacata 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 AGAAGTCCCCAGAACTTTCAAGAAGCGGCGCAAGCATATTTCTTGCTCCCAATG 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 557 tttaattgtctcaaatagaagctgtggaatatagtccacatttccagactccaagctg 616
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 557 TCTTCAATGCCAGTAAGCTGTGTCGGGTGTATTGCTCTGCTGCGGAGTGAACCTG 616
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 617 atttaattgctgtgcgaaccttgcgaagcttcaatccagttgaataatgttcaact 676
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 617 AACTCACTGCTAGGAACATTTGCTCCAGAGCTCAGTCCCTGGAATACTCTGAAATA 676
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 677 tgtgtggaagaaatttccctgaagtaaatlttgaattgtgtcagaagtgtgaataaac 736
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 CATGTGGGAGGAGCTTCTTAAAGAGCAATGAGAGATGCTCAGGCTCTCAAGATGT 736
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 737 tcaatggagcaaatatgtttggaagcgtggaaccccccaagaagaattggaagaattca 796
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 TGAATGGAGGAGATGCAATGGAGTCAAGTACTCTTAAGCACAAGAAACCCGCTGGA 796
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 797 ctaccatcatgaacttagagcgtgtccttataatgtatgtaagctaccataaaggacaa 856
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 AATATCATTGTGAGTAGAGAGACATTAAC-----CTAACCA 838
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 857 acatctccaagaagacaccccccaataacattcaatattgttggcagtgatttatttg 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 837 ACAAGAAGAGGAGCTCTCCCTTAAATTAATTAATGTTTACAGGAGATGCGTCAATG 898
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 917 tttaagtaagaacttftlaaataatttcaaacactccatgcttcaagaacttttttg 976
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 899 TGGCTTCCCGAGATTTCGCCAATGTTTGAAGAACCTTAAATCCCAACAGATGATG 958
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 977 cctgtctaaagaacatactctcgtatgagcacttttgggtcacttgaattgggttc 1036
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 959 AATGGATAAAGACATTATAGCCAGATGAACACTCTTGCGCCACCTTCAGCGTGCAC 1018
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1037 caggaataactgtggagatttccagatcaccccaagatg---tgtcgtatcgcagagta 1093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1019 GGTGAGATGCTGCTGTGTCCCAACACCCCAAGTAGACATCTCAACATGACTCTTA 1078
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1094 agactgcctgttcaagtgttaattactgaagccttttctatccagtt----- 1143
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1079 TTGCGAGGCTGTGCAAGTGGCAGGCTCATGAGGAGACATTCATTAAGGCTGCTTATG 1138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1144 -----tgtactgatctcacttcgaagcgttgtatattatgagctgcagaataaagt 1198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1139 CTTCCCTGCTCTGGATCCACGAGCGGCTATCTGCGTTATATGGGCTGAGACTGAANT 1198
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1199 gactatcaagaatgacattgttctgctcaaaatttgattcctaagttgacccctact 1258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1199 GGATGCTTAAACATCACCCTGTTGGCCCAACAGTTTGACCCCAAGGATGATGAATG 1258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1259 tgataatgcttggcagaagaagct 1283
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1259 CTTTCAAGTCTTAGAAGATTAACCT 1283
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AF038650      1317 bp      mRNA      PRI      25-MAR-1999
LOCUS      Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase
DEFINITION      (C2/4GNT) mRNA, complete cds.
ACCESSION      AF038650
VERSION      AF038650.1 GI:4511880
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 1317)
AUTHORS      Schwienek, T., Nomoto, M., Levery, S.B., Merx, G., van Kessel, A.G.,
Bennett, E.P., Hollingsworth, M.A. and Clausen, H.
TITLE      Control of O-glycan branch formation. Molecular cloning of human
CDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase
forming core 2 and core 4
J. Biol. Chem. 274 (8), 4504-4512 (1999)
99143102
JOURNAL
MEDLINE
REFERENCE      2 (bases 1 to 1317)
AUTHORS      Schwienek, T. and Clausen, H.
TITLE      Direct Submission
JOURNAL      Submitted (28-NOV-1997) Dental School, Faculty of Health Sciences,
Copenhagen University, Noerre Alle 20, Copenhagen 2200, Denmark
FEATURES
source
1..1317
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1317
/gene="C2/4GNT"
1..1317
CDS
/gene="C2/4GNT"
/note="core 2/4-Gnt"
/product="core 2/core 4
beta-1,6-N-acetylglucosaminyltransferase"
/protein_id="ABD21525.1"
/db_xref="GI:4511881"
/translation="MYOMKRIQOLHYLWAGCYMLATVALKLSPELKCDSHDLGLES
RESQSYCRNIIYNFLKLPARKSINSQVTRGQOEVLQAIINLNNEVKKRPPEDTH
YLSLTDCEHFEKAEKFIOPLSKEVEEPFLASVYHKEIEFELRLAYAPONIT
CYAVDEKSPETKEKAEKFIOPLSKEVEEPFLASVYHKEIEFELRLAYAPONIT
PKMYPLNTGCTDEPKISNEMOALKMLNGRNSMESEVPEKRETRMYHFEVBDTL
HLTKRKDDPPYPLNLTFTGNAYIVASRDEVOHYLKNPKSQOILEMYKDYSPDEHLM
TLORARMPGSVPHNPKYDIDSMTISARLYVKOGHGBDIDKGPVAPCSGISHORALICV
YGAIDLNMWLNQNHLLANKRDPKVDNNAQCELYRYAIVGTTEL"
BASE COUNT      359 a      307 c      322 g      329 t
ORIGIN

Query Match      14.1%; Score 191.8; DB 88; Length 1317;
Best Local Similarity 52.5%; Pred. No. 6.4e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

Qy 317 tgaccagtgattgacattatagaagctcaagaagttatgctcaaaagctgtctcaa 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 TCACCAGAGCTGTGAGCATTCAAGGCTGAAGAGAGTTCATACAGTCCCACTGAGCA 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 377 aggaaggaagaaagcttcccaatagcctattcttggttgtccacaagaatgcaatla 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

QY 797 ctaccatcatgaactagacgggtgccttatgaatatgtgaagctaccataagacaa 856
Db 1150 AATATGACTTTGAGTACTGAGAGACACATTAC-----CTAACCA 1191
QY 857 acatcccaagaagacacccccccataaccatcagatatgttgcagtgctatttg 916
Db 1192 ACAAGAAAGAGATCCCTCCCTTATATTACTATTGACAGGGAATCCGACATG 1251
QY 917 tttaagtcaggaattgttaataatatttcaacaactcattgttcaagactttttg 976
Db 1252 TGGCTTCCCGAGATTCTCCACACATGTTGAGAACCCCTTAATCCCAACACTGATG 1311
QY 977 cctgttcaagaacacacactctctgtatgagcaacttgggtcactgtatcgttc 1036
Db 1312 AATGGGTAAAGACACTTATAGCCAGATGACACACTGTGGCCACCTTAGCGTGCAC 1371
QY 1037 caggaaacacctggggagatttccagatcagcccaagatg---tgtatctcagaagta 1093
Db 1372 GGTGGATGCTGGGCTCTGTTCACACACCCCAAGTACGACATCTCAGACATGACTTCTA 1431
QY 1094 agactcgcctgttcaagaagattactatgaagctttttctatcccaagt----- 1143
Db 1432 TTGCCAGGCTGTCAAGTGGCAGGCTCATGAGGAGACATGATAGGCTCTCTTATG 1491
QY 1144 ----tgtactggaactcacccttcgaagcggtgtattatgtgaagctgcagaattaaagt 1198
Db 1492 CTCCTGCTCTGGAATCCACACGCGGCTATCTGCGTTATGGGGCTGGGACCTTGAAAT 1551
QY 1199 ggcctatacaagaatggaacttggttgcataaattgattcgaatggaacctatct 1258
Db 1552 GGATGCTTAAACCATCATCCTGTTGGCCACAAAGTTTGACCCAAAGTATGATATG 1611
QY 1259 tgattaaatgcttggcagaanaagct 1283
Db 1612 CTCCTCAGGCTTAGAATACCT 1636

RESULT 9
AX045249 2216 bp DNA PAT 24-NOV-2000
LOCUS AX045249
DEFINITION Sequence 1 from Patent WO0066727.
ACCESSION AX045249
VERSION AX045249.1 GI:11343799
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2216)
AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
TUMOUR-ASSOCIATED ANTIGEN
JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Source
1..2216
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..426
/feature="unlabeled protein product"
427..1743
/codon_start=1
/protein_id="CAC1734.1"
/db_xref="GI:11343800"
/translat="MWQKRLQLHYLALQCYMLATVALKLSFRKCDSDHIGLES
RESOYCRNIIYNFLKLPARKSINCSVTRGDEAVLOALINLEVKKKEPTDTH
YLSRDCHEFAERKFTOPPLSKKEVEFPLAVSVIHEKINFERLLRAYAONIV
CVHDEKPEFEKAVKALISCFPVFASLKVAVVVAISMERVADINCMDDLOSAY
PKRYLNTGCTDTPFKISNAEMQAKIKMNGNSMESVPRPKHKEKRYKHEVYEDLI
HLTKKRDPPNLTFTGNAYIASRPFVHLNFKSGLIENVKDTISPDHILMA
TLQRARMPGVSNHPRKYDISMTISALVWQGEEDIGAPAPGSGIHORATV
YGAGDLNMMLQNHHLANKPDKYDNLQCLELRYKAIYGTGL"
1744..2216
3'UTR

```

```

BASE COUNT 600 a 499 c 547 g 570 t
ORIGIN
Query Match 14.1%; Score 191.8; DB 9; Length 2216;
Best Local Similarity 52.5%; Pred. No. 6.7e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
QY 317 tgacagtgatgtgtgacattatcaacactcgaagtgatgtcgaagaagctgtctcaa 376
Db 743 TCACCAAGAGCTGTGACACTTCACAGCGTGAAGAGAACTTCATATACGTTCCACAGCA 802
QY 377 aggaagagaaagcttcccaatagcctattcttgggttgcacaaagatgaattatg 436
Db 803 AAGAGAGGTGGAGTCCCTATGACATACCTATGATGATGATGATGATGATGATGATG 862
QY 437 ttgaaggttaccatgcatatcaacagacacatttaccgacattatgac 496
Db 863 TTGAAGGCTACTGCGAGCTGTATGCCCTCAGAACATATCTGTGCATGTGGATG 922
QY 497 gtaagacctgtatccttcaaaagtgtccatgaacaattagctaatgtcttccaaata 556
Db 923 AGAAGTCCCGAGAACTTCAAGAGCGGTCAAGCAATTAATTTCTTCTCCCAAG 982
QY 557 ttctcatgttcccaattagagcgctgtggaataatgccacattccaaactccagctg 616
Db 983 TCTTCATATGACAGTAACTGCTGTGGGTGTTATCTCTCTCTGTGTCACAGGTGCAAGCTG 1042
QY 617 attaaatgtctgtcgaacctctgaagcttcaagctcaatccagtggaataatgtatcaact 676
Db 1043 ACCCTCACTCATGAAAGATGTTGCTCCAGAGCTAGTGGCGGTGAATCTTCTGATA 1102
QY 677 tgggtggcaagatttccctgaagtcaaatlttgatgtgtgaaggttgaanaaac 736
Db 1103 CATGTGGAGGAGCTTCTCATTAAGAGCAATCAGAGATGTCAGGCTCTCAAGATGT 1162
QY 737 tcaatggaacaatatgttggagacggttgaaccccccaacagtaaatgtgaagaatca 796
Db 1163 TGAATGGAGGAGATATGACATGAGATGAGAGTACTCTCTTATGACAAAGAACCCGCTGGA 1222
QY 797 ctaccatcatgaactagacgggtgccttatgaatatgtgaagctaccataagacaa 856
Db 1223 AATATGACTTTGAGTACTGAGAGACACATTAC-----CTAACCA 1264
QY 857 acatcccaagaagacacccccccataaccatcagatatgttggcagtgctatttg 916
Db 1265 ACAAGAAAGAGATCCCTCCCTTATATTACTATTGACAGGGAATCCGATACATG 1324
QY 917 tttaagtcaggaattgttaataatatttcaacaactcattgttcaagactttttg 976
Db 1325 TGGCTTCCCGAGATTCTCCACACATGTTGAGAACCCCTTAATCCCAACACTGATG 1384
QY 977 cctgttcaagaacacactctctgtatgagcaacttgggtcactgtatcgttc 1036
Db 1385 AATGGGTAAAGACACTTATAGCCAGATGACACACTCTGTGGCCACCTTAGCGTGCAC 1444
QY 1037 caggaaacacctggggagatttccagatcagcccaagatg---tgtatctcagaagta 1093
Db 1445 GGTGGATGCTGGGCTCTGTTCACACACCCCAAGTACGACATCTCAGACATGACTTCTA 1504
QY 1094 agactcgcctgttcaagaatggaattactatgaagctttttctatcccaagt----- 1143
Db 1505 TTGCCAGGCTGTCAAGTGGCAGGCTATGAGGAGACATCGATAGGCTGCTCTTATG 1564
QY 1144 ----tgtactggaactcacccttcgaagcggtgtattatgtgaagctgcagaattaaagt 1198
Db 1565 CTCCTGCTCTGGAATCCACACGCGGCTATCTGCGTTATGGGGCTGGGACCTTGAAAT 1624
QY 1199 ggcctatacaagaatggaacttggttgcataaattgattcgaatggaacctatct 1258
Db 1625 GGATGCTTCAAAACCATCATCCTGTTGGCCAAAGTTTGACCCAAAGTATGATATG 1684
QY 1259 tgattaaatgcttggcagaanaagct 1283

```

```

Db      1665 CTCCTCATGCTTGAAGAAATACCT 1709
RESULT 10
AC022480/c
LOCUS   AC022480
DEFINITION
HOMO sapiens chromosome 15 clone RP11-361D15, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
ACCESSION
AC022480
VERSION
AC022480.5 GI:9230840
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 181983)
AUTHORS
Waterston,R.H.
TITLE
The sequence of Homo sapiens clone
JOURNAL
Unpublished
2 (bases 1 to 181983)
REFERENCE
Waterston,R.H.
TITLE
Direct Submission
AUTHORS
Submitted (04-FEB-2000) Genome Sequencing Center, Washington
JOURNAL
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jul 16, 2000 this sequence version replaced gl:7321974.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0361D15
----- Summary Statistics -----
Sequencing vector: M13: 100%
Sequencing vector: plasmid: 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 16456 bases at least Q40
Consensus quality: 17052 bases at least Q30
Consensus quality: 17255 bases at least Q20
Insert size: 16800; agarose-fp
Insert size: 179683; sum-of-contigs
Quality coverage: 3.76 in Q20 bases; agarose-fp
Quality coverage: 3.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1088      1087: contig of 1087 bp in length
* 1188      1187: gap of unknown length
* 2741      2741: contig of 1554 bp in length
* 2742      2841: gap of unknown length
* 2842      4320: contig of 1679 bp in length
* 4521      4620: gap of unknown length
* 4621      6998: contig of 2378 bp in length
* 6999      7098: gap of unknown length
* 7099      10122: contig of 3024 bp in length
* 10123     10232: gap of unknown length
* 10233     14393: contig of 4171 bp in length
* 14394     14493: gap of unknown length
* 14494     18603: contig of 4110 bp in length
* 18604     18703: gap of unknown length
* 18704     22356: contig of 3653 bp in length
* 22357     22456: gap of unknown length
* 22457     26948: contig of 4492 bp in length

```

```

FEATURES
source
* 26949 27048: gap of unknown length
* 27049 31773: contig of 4725 bp in length
* 31774 31873: gap of unknown length
* 31874 36373: contig of 4500 bp in length
* 36374 36473: gap of unknown length
* 36474 41562: contig of 5089 bp in length
* 41563 41662: gap of unknown length
* 41663 46645: contig of 4983 bp in length
* 46646 46745: gap of unknown length
* 46746 53335: contig of 6590 bp in length
* 53336 53435: gap of unknown length
* 53436 60197: contig of 6762 bp in length
* 60198 60297: gap of unknown length
* 60298 68252: contig of 7955 bp in length
* 68253 68352: gap of unknown length
* 68353 75761: contig of 7409 bp in length
* 75762 75861: gap of unknown length
* 75862 84055: contig of 8194 bp in length
* 84056 84155: gap of unknown length
* 84156 94193: contig of 10037 bp in length
* 94194 94292: gap of unknown length
* 94293 109722: contig of 15430 bp in length
* 109723 109822: gap of unknown length
* 109823 125237: contig of 15415 bp in length
* 125238 125337: gap of unknown length
* 125338 142872: contig of 17835 bp in length
* 142873 142972: gap of unknown length
* 142973 159682: contig of 16710 bp in length
* 159683 159783 181983: contig of 22201 bp in length.
Location/Qualifiers
1. 181983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/clone="RP11-361D15"
1. 1087
/note="assembly_name:Contig19"
1188. 2741
/note="assembly_name:Contig24"
2842. 4520
/note="assembly_name:Contig25"
4621. 6998
/note="assembly_name:Contig26"
7099. 10122
/note="assembly_name:Contig27"
10223. 14393
/note="assembly_name:Contig28"
14494. 18603
/note="assembly_name:Contig29"
18704. 22356
/note="assembly_name:Contig30"
22457. 26948
/note="assembly_name:Contig31"
27049. 31773
/note="assembly_name:Contig32"
31874. 36373
/note="assembly_name:Contig33"
36474. 41562
/note="assembly_name:Contig34"
41663. 46645
/note="assembly_name:Contig35"
46746. 53335
/note="assembly_name:Contig36"
53436. 60197
/note="assembly_name:Contig37"
60298. 68252
/note="assembly_name:Contig38"
68353. 75761
/note="assembly_name:Contig39"
75862. 84055
/note="assembly_name:Contig40"
84156. 94192
misc_feature
26949 27048: gap of unknown length
27049 31773: contig of 4725 bp in length
31774 31873: gap of unknown length
31874 36373: contig of 4500 bp in length
36374 36473: gap of unknown length
36474 41562: contig of 5089 bp in length
41563 41662: gap of unknown length
41663 46645: contig of 4983 bp in length
46646 46745: gap of unknown length
46746 53335: contig of 6590 bp in length
53336 53435: gap of unknown length
53436 60197: contig of 6762 bp in length
60198 60297: gap of unknown length
60298 68252: contig of 7955 bp in length
68253 68352: gap of unknown length
68353 75761: contig of 7409 bp in length
75762 75861: gap of unknown length
75862 84055: contig of 8194 bp in length
84056 84155: gap of unknown length
84156 94193: contig of 10037 bp in length
94194 94292: gap of unknown length
94293 109722: contig of 15430 bp in length
109723 109822: gap of unknown length
109823 125237: contig of 15415 bp in length
125238 125337: gap of unknown length
125338 142872: contig of 17835 bp in length
142873 142972: gap of unknown length
142973 159682: contig of 16710 bp in length
159683 159783 181983: contig of 22201 bp in length.

```

```

/misc_feature      /note="assembly_name:Contig41"
                    94293..109722
                    /note="assembly_name:Contig42"
                    109823..125237
                    /note="assembly_name:Contig43"
                    125338..142872
                    /note="assembly_name:Contig44"
                    142973..159682
                    /note="assembly_name:Contig45"
                    clone_end:SP6
                    vector_side:left"
                    159783..181983
misc_feature      /note="assembly_name:Contig46"
                    159783..181983
BASE COUNT      53047 a 36975 c 36896 g 52718 t 2347 others
ORIGIN

```

```

Query Match      14.1%; Score 191.8; DB 67; Length 181983;
Best Local Similarity 52.5%; Pred. No. 8.6e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

Qy 317  tgaccagatgttgacattatcaagactcaagaagttatgctcaaaagctgtctca 376
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86723  TCACCAGAGAGCTGTGAGCACTTCAAGGCTGAAAGAACTTCATACAGTTCCACTGAGCA 86664

Qy 377  aggaagagaagaagctcccaatagactatcttctgtgtccacaagaagatgcaattatg 436
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86663  AAGAAGAGAGTGGAGTTCCCTTTCATACACTATATGATGATCAGAGAGATTGAAACT 86604

Qy 437  ttgaaggcttatccatgctatatatacaacagcaaatattactgcatcattatgctc 496
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86603  TTGAAGAGCTACTGAGACTCTGTATGCCCCCTCAGAAACATATAGTGTGTCATGTGATG 86544

Qy 497  gtaagacaccgataacctcaaaagttgcaagaaattgcaagtgctctccata 556
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86543  AGAAGTCCCGAGAACTTTCAAAGAGGGGTCAAGACATATTTCTGCTCCCAATG 86484

Qy 557  tttaattgtctccaattagaggctgtgaaatgtgccacattccagactccagagctg 616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86483  TCTTATATGACAGTMAAGCTGTGCGGTGTTATGCTCTGCTCAGAGGTGCAAGCTG 86424

Qy 617  attaaattgtctgcagactcttgcgaagcttcaatccagttggaataatgttcaact 676
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86423  ACCTCACTGATGAGAACTTGTCTCCAGAGCTCAGTCCCTGGAATACTCTCTGATA 86364

Qy 677  tgtgtggaagagatttcccttgaaagtcatttgaaattgtgtgcagagttgaaaaaac 736
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86363  CATGTGGAGAGGACTTTCCTATTAAGACCAATGCAGAGATGTCCAGGCTCTCAAGATGT 86304

Qy 727  tcaatgaggaacaatagttttgagagcgtgtgaaaccccccaacagataattgaaagattca 796
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86303  TGAATGGAGAGGATAGCATGTGAGTCAAGTACCTCTTACACACAAAGAACCCGCTGGA 86244

Qy 797  cttaaccatcatgaactttagagcgtgtgccttataatgtgaaagctaacaaagagaa 856
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86243  AATATCACTTGTAGAGTGTGAGAGACACATTACAC-----CTAACCA 86202

Qy 857  aaattccaagaagaagcaccocccataaacattcagataattgttgagcagtgctattttg 916
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86201  ACAAAGAGAGATCTCTCCCTTTATTAATTATATGTATTACAGGAGATCGGATCATTTG 86142

Qy 917  tttaagtaagcatttgttaataatatttcaaacactcactgcttaagaactttttg 976
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86141  TGGCTTCCCGAGATTTCGTCCAAACATGTTTGAAGAACCCCAATCCCAACACAGATTTG 86082

Qy 977  cctgtctaaagaacacactctcctgttgagcacttttgggtctactgttgcaggttc 1036
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86081  AATGGTTAAAGACACTTATACCCAGATGAACACCTCTGGGCCACCTTACGCTGCAC 86022

Qy 1037  caggaaataactgtgagagatttccagatcagcccaagatg---tgtctgactgtcagagta 1093
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86021  GGGGATGGCTGCTGCTGTTCCTCCAAACCAACCCCAAGTACGACATCTCAACATGACTCTTA 85962

```

```

Qy 1094  agactgccttcaagtggaattactatgaagcttttctatccagct----- 1143
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85961  TTGCCAGGCTGTCAAGTGGCGGGGTCAATGAGGAGACATGATAGAGGCTCTCTTATG 85902

Qy 1144  -----tgacttgatctcaccttcgaacgltgttattatgaagctcagaataaagt 1198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85901  CTCCTCGCTCTGTGATTCACACAGCGGGGTACTGCTTATAGGGGCTGGGACTTGAAAT 85842

Qy 1199  ggcattcaagaatgagactgtgtgttataaatttgattcctaagtggaacctact 1258
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85841  GGATGTTTCAAAACCATCCACTGTTGGCCACCAAGTTTGACCAAGGATGATGATAATG 85782

Qy 1259  tgattaaatgcttgcagaagaagct 1283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85781  CTTTCACTGCTCTTACAGATATACCT 85757

```

```

RESULT 11
AF231105      2017 bp      DNA      VRL      26-MAY-2000
LOCUS      Bovine herpesvirus 4 beta-1,6-N-acetylglucosaminyltransferase
DEFINITION      (BORF3-4) gene, complete cds.
ACCESSION      AF231105.1 GI:8096688
VERSION      AF231105.1 GI:8096688
KEYWORDS
SOURCE      Bovine herpesvirus 4.
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae.
ORGANISM      1 (bases 1 to 2017)
            Vanderplassen,A., Markine-Gorlaynoff,N., Lomonte,P., Suzuki,M.,
            Hiroka,N., Yen,U.-C., Bureau,F., Willems,L., Thiry,E., Fukuda,M.
            and Pastoret,P.-P.
            A multipotential beta -1,6-N-acetylglucosaminyl-transferase is
            encoded by bovine herpesvirus type 4
            Proc. Natl. Acad. Sci. U.S.A. 97 (11), 5756-5761 (2000)
            20283898
REFERENCE      2 (bases 1 to 2017)
            Vanderplassen,A., Markine-Gorlaynoff,N., Lomonte,P., Suzuki,M.,
            Hiroka,N., Yen,U.-C., Bureau,F., Willems,L., Thiry,E., Fukuda,M.
            and Pastoret,P.-P.
            Direct Submission
            Submitted (03-FEB-2000) Immunology-Vaccinology, University of
            Liege, Faculty of Veterinary Medicine B43bis, Liege B-4000, Belgium
FEATURES
            Location/Qualifiers
            1..2017
            /organism="Bovine herpesvirus 4"
            /strain="V. test"
            /db_xref="taxon:10385"
            227..1549
            /gene="BORF3-4"
            227..1549
            /gene="BORF3-4"
            /function="forms core 2, core 4, and I branches"
            /note="last ORF at the right end of the L-DNA; similar to
            Homo sapiens core 2
            beta-1,6-N-acetylglucosaminyltransferase, mucin type;
            PBORF3-4"
            /codon_start=1
            /product="beta-1,6-N-acetylglucosaminyltransferase"
            /protein_id="AA072001.1"
            /db_xref="GI:8096689"
            /translation="MKMGKKKLPGRHILNAGCYMLLAVSLRSLRFKQDVNSLD
            LBSRFQSOCHDMILYNSILKPAKRSINSGITRDOQRAVVALDLNLELVKKRPPLT
            DRYVILNTRDCERFRAQRKFIQPLSKBELDPIYASVWEKIEKIFRLRAVAPQ
            NIYCVIVDVKSPEERKAVKAIISCFPNVFNASKLVPAVYASMSRVOADLNCMEDLD
            SSVSKMYLNTGTFDPIKTNAMVNLAKMLGKNSMESEVSESKRMKRYEVD
            TLYPTSKIDPPDNLPMETGNAYVASAFVQHYLDNPKSQIILVEMVKDYSPDEHL
            VATIQCAPMMPGSSVSPHRYHISDMTALAEALRYKQVYHGSDVSMGAPVAPGSIHRAI
            CTYGADLITWLNHLLANKRDPYVDNDVLOCLEDEYLRHKAITYTEL"
BASE COUNT      493 a 504 c 535 g 485 t
ORIGIN

```



Query Match 14.1%; Score 191.6; DB 58; Length 2017;  
 Best Local Similarity 52.5%; Pred. No. 7.4e-36;  
 Matches 518; Conservative 0; Mismatches 429; Indels 39; Gaps 3;

QY 316 atgaccagtgatgtgacattatcagactcaagaattatgtctcaaaagctgtccca 375  
 |||||  
 DB 551 ATACAGGAGGAGCTGTAGCGGCTTCAAGGCCCAAGGAAGTTCATCACTTCCACTGAGT 610  
 |||||  
 QY 376 aagagagaaagaagcttcccaatagactattcttggtgtgtcccaagaatgacat 435  
 |||||  
 DB 611 AAGAGAGAGCTGAGCTTCCCATTTGCCCTTCTGATGAGTGTGTCATGAGAAAGTTAGAAC 670  
 |||||  
 QY 436 gttgaaaggcttatccatgtatatacaaacagcaaatatttactgtatccattatg 495  
 |||||  
 DB 671 TTGAAGCGGCTGCTGGAGCGCTGTATGCCCTCCAGAACATTAAGTGTGTCACAGCTGAT 730  
 |||||  
 QY 496 cgtgaagagacgtatgacattcaaaagtgtgcaatgaacaattagctaaagtgtcccaat 555  
 |||||  
 DB 731 GTGAAGTCCCCAGAGACTTTCANAGAGGCGGTCAAGGCCATTAATTCCTGCCCTCCCAAT 790  
 |||||  
 QY 556 atttcatgtctccaattagagagctgtggaatattgcccacatttccagactcagagct 615  
 |||||  
 DB 791 GTCTCATGAGCCAGTAAGTGTGGTCCGCGTGTATGCTCCTCGTCCAGAGTGAAGCT 850  
 |||||  
 QY 616 gatttaattgtctgtcgaccttctgaagcttcaatccagtgaataatgttataac 675  
 |||||  
 DB 851 GACCTAAGCTGTATGGAAGAGCTTGTCTCAGAGCTCAGTCTATGGAGTACTTACTGAAAT 910  
 |||||  
 QY 676 ttctgtgaggaagatttccctcgaagatcaatttgaattgtgtcagagttgaaaaa 735  
 |||||  
 DB 911 AACTGGGGAGACACTTCCCATTAAGACCAATGCCAGATGATGCTCGCTCCCTCAAGATG 970  
 |||||  
 QY 736 ctcaatgtgagcaaatattgtgagagcgtgtgaaccccccaacagtaatttgaagaat 795  
 |||||  
 DB 971 TTGAAGGGTAAGACAGTATGAGTGTGAGTACCTTCTGAGTCCAAAAAATCGCTGG 1030  
 |||||  
 QY 796 acttaccatcatgaacttgaagagcgtgtgcttattgaatatgtgaagctaccacaataag 855  
 |||||  
 DB 1031 AATTAACCGCTATGAGTGTGAGCAGACACATGTAACCTTACAGCAAGATA----- 1078  
 |||||  
 QY 856 aacatctcaaggaagcaaccccccatatacatcagatatattgtgtgagctgtctt 915  
 |||||  
 DB 1079 -----AAGACCCCTCCCTCGATTAATTACCATGTTCACAGGGAGATGCTATTTT 1129  
 |||||  
 QY 916 gtttaagtaacgatttgttaatatatttcaacaacccaacgltcaagacttttt 975  
 |||||  
 DB 1130 GTGGCCCTCTGAGCCTTTGTCCAAACATGTCTTAGACACACCTTAATCCCAATATCGGTT 1189  
 |||||  
 QY 976 gctgtgtctaaagaacacatactctcctgtatgagcaacttttgggtctactgtatcggt 1035  
 |||||  
 DB 1190 GAATGGGTCAAAAGACATATATAGCCCCGAGAGACACCTCTGGGGCACCTCTGAGCGTGT 1249  
 |||||  
 QY 1036 ccaggaataacctgggagagattccagatcagcccaagatg---tgtctgtctgcagagt 1092  
 |||||  
 DB 1250 CCGTGGATGCTGCTGTCTGTCTTCTAGCCACCCAAAGTATACATCTCAGACATGAGTGGC 1309  
 |||||  
 QY 1093 aagaatcgctgtcaagtggaattactatgaaggtctttc-----tat 1137  
 |||||  
 DB 1310 ATGCCCGAGGCTCGCAAGTGGAGTACCAAGAGGAGATGTTCAGCATGAGGGGGGCGCTTAT 1369  
 |||||  
 QY 1138 cccagtgatgaatgactcagccttcgaagcgtgtgtattatgagctgtgaattaaag 1197  
 |||||  
 DB 1370 GCACCCCTGCTCTGAAATCCATCGAGAGGCGCATCTGCATTTACGGGGCGGGAGCTGTAC 1429  
 |||||  
 QY 1198 tggcttcaaaagatgacatttgggttgcataaatttgaatttctaagtgagccatc 1257  
 |||||  
 DB 1430 TGAATCTCCAGACATCACTCTTGTGCAAAAGATTGACCCGAGGGGTGATGATTAAC 1489  
 |||||  
 QY 1258 ttgattaaatgtctggcagaagaagct 1283  
 |||||  
 DB 1490 GTCTGCGAGTGTGAGAGAGTACTT 1515  
 |||||

RESULT 12  
 MMU68182  
 LOCUS  
 DEFINITION  
 beta-1,6-N-acetylglucosaminyltransferase mRNA, complete cds.  
 068182.1 GI:1763028

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

house mouse,  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

1 (bases 1 to 2393)  
 Magnet, A.D. and Fukuda, M.  
 Expression of large I antigen forming  
 beta-1,6-N-acetylglucosaminyltransferase in various tissues of  
 adult and embryonic mice  
 Unpublished  
 2 (bases 1 to 2393)  
 Magnet, A.D. and Fukuda, M.  
 Direct Submission  
 Submitted (27-AUG-1996) Glycobiology Program, The Burnham  
 Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES  
 source  
 location/Qualifiers  
 1..2393  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 237..1439  
 /codon\_start=1  
 /product="large I antigen-forming  
 beta-1,6-N-acetylglucosaminyltransferase"  
 /protein\_id="AAB39621.1"  
 /db\_xref="GI:1763028"  
 /translation="MPLSVRFVIVATYTMVCVSVYLSFGDQSYOKLNTSDSVRL  
 SOYCSFIDKSRFLRNKLMHEKSCFETVYOSVITAPLSOEGDPLAVYVTH  
 HNDFEALRLFLFMPONITGVHDEKATREFGAYBOLVSCPYNFLASKPEPVYVG  
 GISHLQDLNCLIDLSTSEVPWYVATTCQDPLPTNKIVGLGSKNTPSYL  
 PPAALIGRTYVHREHLSKEISYVIRTPALKPPEPNLTVFSASVALSREPANVL  
 RDPRAVDLLHMSKDTPSDHFVWTLNRIGVGSPPNSMTGNLAVVMDEAARG  
 GCGHYVHGICIVGNGLQWLINSQSLFANKFELNTPVIVECLELRRLRRLINQSES  
 ALQPSWYF"

BASE COUNT 610 a 593 c 583 g 601 t  
 ORIGIN

Query Match 14.0%; Score 191.2; DB 94; Length 2393;  
 Best Local Similarity 52.6%; Pred. No. 9.3e-36;  
 Matches 502; Conservative 0; Mismatches 428; Indels 24; Gaps 3;

QY 337 tatcagacttaagaaggtatagctcaaaagcttgcataaaggaggaagaagcttccca 396  
 |||||  
 DB 459 TATGTGACCCCAAGCCACTATATCACTGCCCTTATCTCAGAGAGAGGTGATTTCT 518  
 |||||  
 QY 397 atagccattcttctgtgtccacaagaatgcaattatggttgaagaagcttatccagt 456  
 |||||  
 DB 519 TTGGGGTATGATGTGATCATCATCAATTGTGACACTTTCGACAGGCTCTTCAGGGCA 578  
 |||||  
 QY 457 atatacaaccagacaataatttactgcatcatttgttgcgttaagggccccgataacctc 516  
 |||||  
 DB 579 ATCTTCATGCTCCCAAAATATCTATCTGTCTCAGTGTGAAAAGCAACAGCTGAATT 638  
 |||||  
 QY 517 aaagtgtgcaatgaacaattatgaagtgctctcccaatatcttcaatgtctccaaat 576  
 |||||  
 DB 639 AAGAGTCCGTGGAACAGTATAGTGTGCTTCCCAATATGCTTCTGCTCTTAAGATG 698  
 |||||  
 QY 577 gaggctgtggaatagccacattccagactcagagctgtatttaattgtctgtcgagc 636  
 |||||  
 DB 699 GAGCGGTGTGTATGTTGGAATCTCCGGCTCCAGGCTGACCTTAAGTGTCAAAAGAC 758  
 |||||  
 QY 637 ctctcgaagcttcaatccagtggaataatgttcaacttgtgtggcagaagatttccc 696  
 |||||  
 DB 759 CTTTCACCTCCGAGGTCCCTTGGAATGACCATCAACACTCTGTGACAAAGACTTCCC 818  
 |||||



repeat_region	/note="L1 repeat: matches 2781. 2830 of consensus"	11701. 11793
repeat_region	/note="MER82 repeat: matches 1. .96 of consensus"	11794. 12114
repeat_region	/note="LIME2 repeat: matches 5315. .5644 of consensus"	12157. 12221
repeat_region	/note="L1PA11 repeat: matches 6095. .6164 of consensus"	12246. 12683
repeat_region	/note="L1M4 repeat: matches 2774. .3210 of consensus"	12687. 12824
repeat_region	/note="MSTA repeat: matches 238. .370 of consensus"	12863. 12933
repeat_region	/note="L1M4 repeat: matches 2691. .2758 of consensus"	12934. 13237
repeat_region	/note="AlusX repeat: matches 1. .305 of consensus"	13238. 13293
repeat_region	/note="L1M4 repeat: matches 2635. .2691 of consensus"	13346. 13830
repeat_region	/note="L1M6C repeat: matches 361. .853 of consensus"	14517. 14815
repeat_region	/note="Aluub repeat: matches 1. .297 of consensus"	14909. 15069
repeat_region	/note="Aluub repeat: matches 143. .297 of consensus"	15977. 16273
repeat_region	/note="AlusX repeat: matches 1. .297 of consensus"	16378. 16502
repeat_region	/note="FLM4C repeat: matches 2. .125 of consensus"	16959. 17263
repeat_region	/note="AlusX repeat: matches 1. .305 of consensus"	17264. 17287
repeat_region	/note="12 copies 2 mer ta 100% conserved"	17464. 17698
repeat_region	/note="L1M87 repeat: matches 5928. .6169 of consensus"	18053. 18080
repeat_region	/note="L2 repeat: matches 2662. .2689 of consensus"	18094. 18272
repeat_region	/note="FAM repeat: matches 1. .175 of consensus"	18355. 18547
repeat_region	/note="MER20 repeat: matches 26. .218 of consensus"	18714. 19001
repeat_region	/note="L1M82 repeat: matches 5850. .6142 of consensus"	19038. 19344
repeat_region	/note="AlusX repeat: matches 5. .309 of consensus"	19578. 19615
repeat_region	/note="ERNA-Gln-CAG repeat: matches 1. .38 of consensus"	20357. 20715
repeat_region	/note="MER46B repeat: matches 2. .236 of consensus"	20799. 21097
repeat_region	/note="AlusX repeat: matches 1. .299 of consensus"	21832. 22018
repeat_region	/note="MIR repeat: matches 21. .220 of consensus"	22167. 22496
repeat_region	/note="MER58B repeat: matches 1. .338 of consensus"	22598. 23024
repeat_region	/note="L1M6 repeat: matches 5862. .6300 of consensus"	23081. 23278
repeat_region	/note="L2 repeat: matches 2394. .2570 of consensus"	23279. 23404
repeat_region	/note="Aluub repeat: matches 1. .130 of consensus"	23405. 23716
repeat_region	/note="AlusP repeat: matches 1. .313 of consensus"	23717. 23735
repeat_region	/note="Aluub repeat: matches 130. .148 of consensus"	23736. 24036
repeat_region	/note="AlusS repeat: matches 1. .301 of consensus"	24037. 24200
repeat_region	/note="Aluub repeat: matches 148. .301 of consensus"	24201. 24214
repeat_region	/note="L2 repeat: matches 2336. .2394 of consensus"	26110. 26393
repeat_region	/note="AlusX repeat: matches 1. .284 of consensus"	26508. 26900
repeat_region	/note="L1M42 repeat: matches 5893. .6306 of consensus"	

	repeat_region	/note="Alusx repeat: matches 1..304 of consensus"	26934..27236
Db	repeat_region	/note="Alusx repeat: matches 1..298 of consensus"	27466..27765
	repeat_region	/note="Alusx repeat: matches 1..298 of consensus"	28230..28265
	repeat_region	/note="LIMC/D repeat: matches 5761..5796 of consensus"	28281..28552
	repeat_region	/note="Alubx repeat: matches 42..312 of consensus"	28553..28845
	repeat_region	/note="Alux repeat: matches 1..293 of consensus"	29137..29431
	repeat_region	/note="Alusx repeat: matches 1..296 of consensus"	30333..30639
	repeat_region	/note="Alusx repeat: matches 1..308 of consensus"	30647..30958
	repeat_region	/note="Alusx repeat: matches 1..313 of consensus"	30973..31270
	repeat_region	/note="Alusx repeat: matches 1..305 of consensus"	31350..31482
	repeat_region	/note="FLAN_A repeat: matches 1..133 of consensus"	32542..32606
	repeat_region	/note="L2 repeat: matches 2679..2746 of consensus"	32885..33161
	repeat_region	/note="Alusx repeat: matches 10..228 of consensus"	34121..34431
Query Match		13.8% ; Score 187.4; DB 90; Length 87636;	
Best Local Similarity		55.4%; Pred. No. 9.5e-35;	
Matches 387; Conservative		0; Mismatches 306; Indexes 6; Gaps	
Qy	345	tctaagaagttatgcccacaagaactgtctccaaaggagaagaacgttcccaataccta	404
Db	40131	TGCAAGCGACTATGTACGAACAACACTCTCTGAAAGAAGAGCGTGGTTCCCTTAGCTTA	40190
Qy	405	tctcttggttcaccaaaagtatcaattatggttgaaggcttatccatgcytatatcaa	464
Db	40191	CACAGTGACCATTCCAAAAGACTGGGCCTTTTGAAGGCTCTTCAGGGCGAATTATAA	40250
Qy	465	cagacaaatattatgatcatcoatatgatagttaaggcacctbaatcctccaagttgc	524
Db	40251	GCCCCAAAATGTCTACTGTTGTGCACCTGGATAGAAAGCGACGATGCTTTAAAGTGC	40310
Qy	525	catgaacaattagccaagtgtcttctccaatatttctatgtcttccaatagaagtcgt	584
Db	40311	AGTAAACAGTTACTCAGGCTGCCCAATCTTTCTGGCTTCCAAGAAGAGTGGT	40370
Qy	585	ggaaatagccccaatcttccagactccaagctatgaattatgtctgtcgagaccttcga	644
Db	40371	TGTTATAGGGGGATCTCCAGCGCTCCAGGCTAACCTGAATGCTCGAAGACCTTGTCG	40430
Qy	645	gtctcaatccaagtggaatatgtltaaccaacttgttggccaagatlttccctcgaagtc	704
Db	40431	CTCTGAGTTCCCTGGAGAGTATGTCATCAACACGTGGGGCAAGCTTCCCCTGAANAAC	40490
Qy	705	aaatttgaattgtgtccagagttgaaaaaacatgtgagcaataatgttggagagcgt	764
Db	40491	CAACAGGGAATATGTTATGATATCTGAAAGGATTTAAAGGAAAAATATCACCCCGAGT	40550
Qy	765	gaagcccccaaaagtaaatgtgaagatcttaccatccatctgaacttagaaggtgccc	824
Db	40551	GCTGCTCTCCAGCACCGCTTTTGGAGCGACTAATATCGTCCACCA-----AGAACTGT	40604
Qy	825	ttatgaatatgtgaagctaccaataagaacaacatctccaaggaagcaccoccccataa	884
Db	40605	AAACACCAAATATCTCATGCTGATTTAAACACACAAATTTAAACCTCTCCTCTCATGA	40664
Qy	885	catcagaatattgttggcagtgcttatatttftttaagtaaagaacttgttaaatat	944
Db	40665	CATGCTGATTTACTTTGGACAGGCTTACGTGGCTTCAACAAAGGACTTTGCTAATTGCT	40724
Qy	945	tttcaaacaccatcgtltaagaactttttgctgtgtltaagaacataatctctcga	1004
Db	40725	CCTCAAGACCAAGCTGACCTGACTTACTACTCTCTGCTCAAGAGCAACCTCAGCCCGGA	40784



Db 102271 CACAGTACACATCCAAAGAGCTTGGACATTTTGGAGAGCTTTCAGGGCGATTATAT 102330  
 Qy 465 ccagacacatttaccgacacattatgacgaagagcccttacccttcaagtgc 524  
 Db 102331 GCGCCAAATCTCTACTGTGTGACCTGGATCAGACAGCGAGATGCTTTTAAAGTGC 102390  
 Qy 525 catgaacatttagcctaagtctctcccaatttccatttcttcccaattagaagcgt 584  
 Db 102391 AGTGAACAGTACTACACTGCTCTCCCAATGCTTTCTGCTCCCAAGAGAGTGGGT 102450  
 Qy 565 ggaatagcccaatttcccaattccagccgatttaatttgcctgcgcagccctcga 644  
 Db 102451 TGTCTATGGGGGATCTCAGGCTCCAGCTCCAGCTGAACTGCTGGAAGAGCTTGCC 102510  
 Qy 645 gtcttcaatcagtggaatatgtatcaattgtgtggaagatttccctcgaagtc 704  
 Db 102511 CTCTGAATTCCTGGAGTATGTGATCAACCTGGGGGACAGACTTCCCTGAAAC 102570  
 Qy 705 aaatttgattgtgtcagagtggaataacccaatgaagcaatattgttgagagcgt 764  
 Db 102571 CAAACAGGAAATAGTTCAGTATCTGAAAGGATTTAAAGGAAATATACACCCCGAGT 102630  
 Qy 765 gaaccccccaacagtaattggaagattcaactacatgaactgaagcaggtgc 824  
 Db 102631 GCTGCTCTCTACACAGCTGTGGACGACATAATACCTCCACCA-----AGAACTGTT 102684  
 Qy 825 ttaagaatatggaagtaaccaataagaacaacatctccagaagaagcaccctccata 884  
 Db 102685 AAACCAACAAATTCCTCTAGTATCTGAAACAAATTAATAAAGCTCTCTCTCATGA 102744  
 Qy 885 gattcagatattgttgcagtgctatttgttgaagcaagcattgttgaatatat 944  
 Db 102745 CATGGTATTTACTTGTGACAGGCTTACGTGCTCTCAAGAGGACTTGTCTACTTGT 102804  
 Qy 945 ttccacaactccatcagtcgaagacttttgcctggtcctaagaacacatactctcga 1004  
 Db 102805 CCTCAAGACCAAGCTGCGACTTACTCTCTGCTGCAAGGACACTCAGACCCCGCA 102864  
 Qy 1005 tgaagacttttggctacacttgatcggttcagaagt 1043  
 Db 102865 CCAACATTTCTGGTGACACTCAGAGGATTTCCCGTAT 102903

RESULT 15  
 AC021626/c  
 LOCUS AC021626  
 DEFINITION Homo sapiens clone RP11-391I2, WORKING DRAFT SEQUENCE, 28 unordered pieces  
 AC021626  
 AC021626.3 GI:7229839  
 VERSION HTG: HTGS\_PHASE1, HTGS\_DRAFT.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 196788)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens, clone RP11-391I2  
 Unpublished  
 2 (bases 1 to 196788)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Balwit,J., Barna,N., Beckert,R., Beda,F., Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Castle,A., Chospel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Deaveliano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardina,S., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,D., Menes,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N., Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

# TITLE JOURNAL COMMENT

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliou,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A., and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 12, 2000 this sequence version replaced g1:6978227.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: 391.I-2  
 Center clone name: 15883

----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 181094 bases at least Q40  
 Consensus quality: 189581 bases at least Q30  
 Consensus quality: 192532 bases at least Q20  
 Insert size: 208000; agarose-fp  
 Insert size: 194088; sum-of-contigs  
 Quality coverage: 3.7 in Q20 bases; agarose-fp  
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 28 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1281: contig of 1281 bp in length  
 \* 1282 1381: gap of 100 bp  
 \* 1382 2889: contig of 1518 bp in length  
 \* 2890 2999: gap of 100 bp  
 \* 3000 5681: contig of 2682 bp in length  
 \* 5682 5781: gap of 100 bp  
 \* 5782 8351: contig of 2570 bp in length  
 \* 8352 8451: gap of 100 bp  
 \* 8452 10639: contig of 2188 bp in length  
 \* 10640 10739: gap of 100 bp  
 \* 10740 13655: contig of 2916 bp in length  
 \* 13656 13755: gap of 100 bp  
 \* 13756 15927: contig of 2172 bp in length  
 \* 15928 16027: gap of 100 bp  
 \* 16028 19028: contig of 3001 bp in length  
 \* 19029 19128: gap of 100 bp  
 \* 19129 22668: contig of 3541 bp in length  
 \* 22670 22769: gap of 100 bp  
 \* 22770 26791: contig of 4022 bp in length  
 \* 26792 26891: gap of 100 bp  
 \* 26892 30845: contig of 3954 bp in length  
 \* 30846 30945: gap of 100 bp  
 \* 30946 35594: contig of 4649 bp in length  
 \* 35595 35694: gap of 100 bp  
 \* 35695 40148: contig of 4454 bp in length  
 \* 40149 40248: gap of 100 bp  
 \* 40249 45042: contig of 4794 bp in length  
 \* 45043 45142: gap of 100 bp  
 \* 45143 49477: contig of 4335 bp in length  
 \* 49478 49577: gap of 100 bp  
 \* 49578 54518: contig of 4941 bp in length  
 \* 54519 54618: gap of 100 bp  
 \* 54619 61982: contig of 7364 bp in length  
 \* 61983 62082: gap of 100 bp

misc_feature	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="RP11-391I12" /clone_id="RP11-391I12.Human Male BAC 1..1281 /note="Assembly-fragment" 1382..2889 /note="assembly-fragment" 3000..5681 /note="assembly-fragment" 5782..8351 /note="assembly-fragment" 8452..10633 /note="assembly-fragment" 10740..11655 /note="assembly-fragment" 13756..15927 /note="assembly-fragment clone_end:17 vector_side:right" 16028..19028 /note="assembly-fragment" 19129..22669 /note="assembly-fragment" 22770..26791 /note="assembly-fragment" 26892..30845 /note="assembly-fragment" 30946..35594 /note="assembly-fragment" 35695..40148 /note="assembly-fragment" 40249..45042 /note="assembly-fragment" 45143..49477 /note="assembly-fragment" 49578..54518 /note="assembly-fragment" 54619..61362 /note="assembly-fragment" 62083..67811 /note="assembly-fragment" 67912..74070 /note="assembly-fragment" 74171..80685 /note="assembly-fragment" 80786..89056 /note="assembly-fragment" 89157..99104 /note="assembly-fragment"
--------------	---

Query Match	13.8%	Score 187.4	DB 66	Length 196788
Best Local Similarity	55.4%	Pred. No. 1e-34		
Matches	387	Conservative	0	Mismatches 305; Indels 6; Gaps 1
QY 345	tctaagaggtatgctccaagaagctgtctccaaggagagaagaaagcttcccaatgcta	404		
DB 78350	TCGAAGCCACTCTGTGAAAGAAACACTCTCTGAAAGAAAGCGCTGGTTCCCTTTAGCTTA	78291		
QY 405	ttccttggtgtgccacaagaatgaactcttggttgaagagctacacatgcatatacga	464		
DB 78290	CACACTGACCAATCCCAAAAGACTTGGGCACCTTTAGAGAGCTCTTCAGGGCATTTAAT	78251		
QY 465	ccagacacataattactgcatccattatgatcgtgaagagcaacctgataccttcaagttgc	524		
DB 78230	GCCCCAAATGCTCTGCTGTGCTACCTGGATCAGAGCGAGGATGCTTTAAAGGTGC	78171		
QY 525	catgaacaattagctaagttgcttcccaataatttcccttgctccaattagaagcgt	584		
DB 78170	ACTGAAACGTTACTACAGCTGCTTCCCAAAAGCTTTCTGGCTTCCAAAGAAAGATCGT	78111		
QY 585	ggaatatgcccacattccacagatccagagctatttaattgctgtgcaccttgaa	644		
DB 78110	TCTATATGGGGGATCTCCAGCTCCAGCGTCAACTGAACTCCGTGAAGACCTTGTGC	78051		
QY 645	gtctccaatccagttggaatatgataactgtgtgagcaagtttcccttaagtc	704		
DB 78050	CTCTGAAGTTCCTTGGAAAGTATGATCATCAACCTCGCGGCAAGACTTCCCTTAAGAC	77991		
QY 705	aaatttgaattcgtgtccagagttgaaaaaacatcagtgagcaaalatgttgaagacgt	764		
DB 77990	CAACAGGGAATAGTTCACTATCTGAAGGGATTTAAAGGAAAAATATCACCCCGGACT	77931		
QY 765	gaaacccccaacagtaaatgtgaagaagatctacatctatgaacttaagcaggtgcc	824		
DB 77930	GCTGCTCTGTGCCACGCTGTGGACGAGCTAAATATGCTCCCA-----GAACTGTT	77877		
QY 825	ttaatgaatgtgaagcttaaccaataaagagacaacatctccaagyaagcaccgcccataa	884		
DB 77876	AAACCACAAAAATTCCTACGTATTAATAACAACAATAATTAATAAATCCCTCCCTCAGTA	77817		
QY 885	catcagaataattgtgtgcagtgctatttgttttaagtcagaacattgttaataatat	944		
DB 77816	CATGTGATTTACTTTGGACAGCGCTACGTGCTTCCAAAGGACTTTGCTAACTTGT	77757		
QY 945	tttcaacaatccatcgttcaagaacttttgcgtgtctaaagacaatctctctga	1004		
DB 77756	CCTCAAGACCAAGCTCGCACTGACTTACTCTCCGTGCTCAAGCAACACTTAAGCCCGA	77697		
QY 1005	tgagcaatttgggtctacctgtatccggttccaggaat	1043		
DB 77696	CGAATCTTCTGGTGACACTCAACAGGATTCCTCGGTAT	77658		

Fri Aug 3 09:38:03 2001

Job time: 6969 sec

us-09-645-292-1.rge

---

Page 19

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:49:53 ; Search time 42.02 Seconds  
(without alignments)  
1306.677 Million cell updates/sec

Title: US-09-645-192-2\_COPY\_39\_453  
Perfect score: 2194  
Sequence: 1 POKDIYLVYSLSTSPFVNN.....DWITLPSKLEFMDRLTTTS 415

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL-16:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mnc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-unclassified:\*
- 13: sp-vertebrate:\*
- 14: sp-virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2194	100.0	453	4 Q9P109	Q9P109 homo sapien
2	862	39.3	438	4 Q9S395	Q9S395 homo sapien
3	854.5	38.7	428	11 Q64165	Q64165 rattus norv
4	848	38.0	440	14 Q91KX2	Q91KX2 bovine hepr
5	833	38.0	428	11 Q35981	Q35981 mus musculu
6	508.5	23.2	314	4 Q9HCV8	Q9HCV8 homo sapien
7	321.5	14.7	445	5 Q19730	Q19730 caenorhabdi
8	313.5	14.3	425	5 Q02234	Q02234 caenorhabdi
9	296	13.5	401	5 Q02314	Q02314 caenorhabdi
10	291	13.1	454	5 Q19729	Q19729 caenorhabdi
11	288.5	13.1	472	5 Q20406	Q20406 caenorhabdi
12	265	12.1	402	5 Q02315	Q02315 caenorhabdi
13	245	11.2	459	5 Q02309	Q02309 caenorhabdi
14	231	10.3	876	5 Q9W034	Q9W034 drosophila
15	226	10.3	454	5 Q45866	Q45866 caenorhabdi
16	224.5	10.2	864	11 Q9EP10	Q9EP10 rattus norv
17	219.5	10.0	865	4 Q9H1B5	Q9H1B5 homo sapien
18	218	10.0	470	5 Q16776	Q16776 caenorhabdi
19	218	9.9	489	5 Q21796	Q21796 caenorhabdi

20	214.5	9.8	616	11 Q9EP10	Q9EP10 mus musculu
21	213.5	9.7	486	5 Q93643	Q93643 caenorhabdi
22	213.5	9.7	821	11 Q9EP11	Q9EP11 rattus norv
23	213.5	9.7	827	4 Q9H1B6	Q9H1B6 mus musculu
24	211.5	9.6	789	11 Q9EP11	Q9EP11 mus musculu
25	196.5	9.0	472	5 Q76634	Q76634 caenorhabdi
26	196	8.9	753	5 Q02312	Q02312 caenorhabdi
27	194.5	8.9	448	5 Q9Z0X7	Q9Z0X7 arabidopsis
28	194.5	8.9	513	5 Q9V3P0	Q9V3P0 caenorhabdi
29	188	8.6	478	5 Q9X112	Q9X112 caenorhabdi
30	187	8.5	447	5 Q9X171	Q9X171 arabidopsis
31	182.5	8.3	513	5 Q9XTM4	Q9XTM4 caenorhabdi
32	181	8.2	424	5 Q9LE60	Q9LE60 arabidopsis
33	178	8.1	367	5 Q22481	Q22481 caenorhabdi
34	178	8.1	467	5 Q9U3H9	Q9U3H9 caenorhabdi
35	175.5	8.0	406	10 Q9LNS5	Q9LNS5 arabidopsis
36	172.5	7.9	378	10 Q9S569	Q9S569 arabidopsis
37	166.5	7.6	434	10 Q9LF00	Q9LF00 arabidopsis
38	151	6.9	447	10 Q9FLD7	Q9FLD7 arabidopsis
39	132	6.0	630	10 Q80927	Q80927 arabidopsis
40	131	6.0	384	10 Q9S253	Q9S253 arabidopsis
41	123.5	5.6	371	5 Q20081	Q20081 caenorhabdi
42	119	5.4	316	2 Q56048	Q56048 streptococ
43	111	5.1	1037	5 Q9NHF7	Q9NHF7 drosophila
44	108	4.9	702	1 Q92829	Q92829 archaebact
45	107	4.9	1441	14 Q9QKFO	Q9QKFO californa

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	453 AA.
Q9P109	Q9P109			
ID	Q9P109			
AC	Q9P109			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	CORE 2 BETA-1,6-N-ACETYLGLUCOSAMINYLTANSFERASE 3.			
GN	C6GNT3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20219156; PubMed=10753916;			
RA	Schwentek T., Yeh J.C., Levery S.B., Keck B., Merx G.,			
RA	van Kessel A.G., Fukuda M., Clausen H.;			
RT	"Control of O-glycan branch formation. Molecular cloning and			
RT	characterization of a novel thymus-associated core 2 beta1,6-N-			
RT	acetylglucosaminyltransferase.";			
RL	J. Biol. Chem. 275:11106-11113(2000).			
DR	EMBL; AF132035; AAF63156.1; -			
DR	InterPro; IPR003406; -			
DR	Pfam; PF02485; Branch; 1.			
KW	Transferase; Glycosyltransferase.			
SQ	SEQUENCE 453 AA; 53052 MW; B43794D427FA1CA CRC64;			
Query Match	100.0%;	Score 2194;	DB 4;	Length 453;
Best Local Similarity	100.0%;	Pred. No. 7,6e-164;		
Matches	415; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	1 POKDIYLVYSLSTSPFVNNRYTVKDEVRYVNCSCIYEDEPLICKSEIRRDITDL 60			
DB	39 POKDIYLVYSLSTSPFVNNRYTVKDEVRYVNCSCIYEDEPLICKSEIRRDITDL 98			
QY	61 EDDVVAATSCDIYQIRGTAOKIVSKSEKSFPIASLVVHKRAIWERLIHAIYQHN 120			
DB	99 EDDVVAATSCDIYQIRGTAOKIVSKSEKSFPIASLVVHKRAIWERLIHAIYQHN 158			
QY	121 IYCIHYDRKAPDTEFVANNMLAKCFSNIFIAKLEAVEYAHISRLQADNCLSDLKSSI 180			



```

Db 159 IYCHIDRKAPDTFTKAMNNLAKCSNIFIASKLEAVEYAHISRLQADINCLSDLKSSI 218
QY 181 QMKYVNLGCGDPLKSNFELVSELKKNANLLETVKPPNSKLEFTHHELRVPY 240
Db 219 QMKYVNLGCGDPLKSNFELVSELKKNANLLETVKPPNSKLEFTHHELRVPY 278
QY 241 VKIPRTINSKEAPPHNIQIFVGSAFYVLSQAFVKYIFNNSIVQDFPAMSKDTPSDEHF 300
Db 279 VKIPRTINSKEAPPHNIQIFVGSAFYVLSQAFVKYIFNNSIVQDFPAMSKDTPSDEHF 338
QY 301 WALLIRPQIPGELSRACQVSDQSKTRLYKNNYIEGFFPSCGSHLRVCITYAAL 360
Db 339 WALLIRPQIPGELSRACQVSDQSKTRLYKNNYIEGFFPSCGSHLRVCITYAAL 398
QY 361 RWLIKDHWFANKFDSKVPDILIKCLAEKLEEQORDMITLPESEKLEMDRNLTTTS 415
Db 399 RWLIKDHWFANKFDSKVPDILIKCLAEKLEEQORDMITLPESEKLEMDRNLTTTS 453

RESULT 2
ID 095395 PRELIMINARY; PRT: 438 AA.
AC 095395;
DT 01-MAY-1999 (TRENBLREL. 10, Created)
DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)
DE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
GN C2/4GNT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99115671; PubMed=9915862;
RA Yeh J.C., Ong E., Fukuda M.;
RT "Molecular cloning and expression of a novel beta-1, 6-N-
RT acetylglucosaminyltransferase that forms core 2, core 4, and I
RT branches.";
RL J. Biol. Chem. 274:3215-3221(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99143102; PubMed=9988682;
RA Schwientek T., Nomoto M., Levery S.B., Merkx G., van Kessel A.G.,
RA Bennett E.P., Hollingsworth M.A., Clausen H.;
RT "Control of O-glycan branch formation. Molecular cloning of human cDNA
RT encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core
RT 2 and core 4.";
RL J. Biol. Chem. 274:4504-4512(1999).
DR EMBL; AF102542; AAD10824.1;
DR EMBL; AF038650; AAD21525.1;
DR InterPro; IPR003406;
DR Pfam; PF02485; Branch; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 438 AA; 50863 MW; 1FF0A7B451C88407 CRC64;

Query Match 39.3%; Score 862; DB 4; Length 438;
Best Local Similarity 44.5%; Pred. No. 1,5e-59;
Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps 8;

```

```

Db 166 EKSPTEKAVKAILISCPNVIASKLVRVYASWSRVQADNCEMLLOSSVPMTYLN 225
QY 188 LOGDPEPLKSNFELVSELKKNANLLETVKPPNSKLEFTHHELRVPY 247
Db 226 TCGDTFPKSNAMVQALKMLNGNSMESVPRPKHKEIRMKYHDEYVR---DTLHL---T 279
QY 248 NISKAPPHNIQIFVGSAFYVLSQAFVKYIFNNSIVQDFPAMSKDTPSDEHFATLIRV 307
Db 280 NKKKDPPEYNLMFTGNAYIVASRDEVOHLKNSQDLLEWVDITSPDEHLMATLQRA 339
QY 308 PCIPGEI-SRSADQVSDQSKTRLYKNNYIEGFF-----YPCGSHLSRVCITYAALR 361
Db 340 RMPGSSVPRHKEKYDLSDMTSTARLYKMGCHGDDIDKGAAPACGSHIRALCYGAGDIN 399
QY 362 WLKDHWFANKFDSKVPDILIKCLAEK 390
Db 400 WMLQNHLLANKFDPKVDNALQCLEYL 428

RESULT 3
ID 064165 PRELIMINARY; PRT: 428 AA.
AC 064165;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)
DE ENZYMIC GLYCOSYLATION-REGULATING GENE PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RX MEDLINE=96013636; PubMed=7560067;
RA Nishio Y., Warren C.E., Buczek-Thomas J.A., Rulfs J., Koya D.,
RA Aiello L.P., Feener E.P., Miller T.B., Jr., Dennis J.W., King G.L.;
RT "Identification and characterization of a gene regulating enzymatic
RT glycosylation which is induced by diabetes and hyperglycemia
RT specifically in rat cardiac tissue.";
RL J. Clin. Invest. 96:1759-1767(1995).
DR EMBL; S79797; AAB35697.2;
DR InterPro; IPR003406;
DR Pfam; PF02485; Branch; 1.
SQ SEQUENCE 428 AA; 49826 MW; AE25A18172897AAB CRC64;

Query Match 38.9%; Score 854.5; DB 11; Length 428;
Best Local Similarity 42.1%; Pred. No. 5,5e-59;
Matches 177; Conservative 71; Mismatches 133; Indels 39; Gaps 10;

```

```

Db 302 NKNIQEMENADTSPDEFMATIQRIPEVPSLPSSEHKYDLSDMNAVARFKWQYEEG 361
Qy 339 FF-----YPSCTGSHLSVCYIAGAEIRMLIKDGHMANKFDSVDPILKCLAEKLEEQ 393
Db 362 DVSNGADYPPCSGVHVSVCYFVGVDLSMMLRKHHFFANKFDMDVDPAQLCLAEHLRHK 421

RESULT 4
091KZ2 PRELIMINARY; PRT: 440 AA.
AC 091KZ2:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
GN BOREF3-4.
OS Bovine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=10385;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V. TEST:
RX MEDLINE=2023898; PubMed=10811864;
RA Vanderplisschen A., Markine-Gorlaynoff N., Lomonte P., Suzuki M.,
RA Hiraoka N., Yen J.-C., Bureau F., Willems L., Thiry E., Fukuda M.,
RA Pastoret P.-P.;
RT "A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded
RT by Bovine herpesvirus type 4."
RL Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761(2000).
DR EMBL: AF231105; AAF72001.1;
DR InterPro: IPR003406;
DR Pfam: PF02485; Branch: 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 440 AA; 50701 MW; 71B9C3B6BA949BD CRC64;

Query Match 38.7%; Score 848; DB 14; Length 440;
Best Local Similarity 44.0%; Pred. No. 1.8e-58;
Matches 168; Conservative 72; Mismatches 124; Indels 18; Gaps 6;

Qy 22 YTHKDEVREYVNCSTY--BOEPL--EIGKSLERRDDIDLEDVYAMTSDCDIYQT 77
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 YNSKLPKRSINCSGITRGDEAVQALDNLNLEVKKR--PLDTGYLNTTRDCERFKA 118
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 78 LRGYQKIVSKSEKSFPAYSGLVHKDAIMVERLIHAIYGNHNYCIHYDKADTERVA 137
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 ORKFIQPLSKBELDEPLAISMVYHEKTEFERLLRAYAPONTICVHVDVKSPEYKEA 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 138 MNNLAKCFNSNFIASKLEAVEYAHISRLQADLNCSDLKSSIQMKYVYNLCGDFPLKS 197
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 VKAISCSPNVEASKLVPYVYASMSRVQADLNCMEDLQSSQMKYLLNCGDFPIKT 238
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 198 NFEVSELKINGANMLTVPKPNKSLERFTYHHLRVPYEVYKLPRTNISKAPPHN 257
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 NAEWLAKMKLKGKMSSEVPSESKNRMKRYREVDITLPTSKI-----ADPPDN 291
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 258 IQIEGSAVFYLSQAFVYIFNNSIVQDFPMSKDTYSPDEHFWATLIRVGPGEI-SR 316
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 LPMFTGNAYFAASRAFGVHLDNPKSQILVEMVYKDTYSPDEHFWATLQRAWMGSPVSH 351
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 317 SAQVSDLSQSTRLYKMYIGSFF-----YPSCTGSHLSVCYIAGAEIRMLIKDGHMFA 371
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 PKYHISDMTALARKVQWYHGGDVSMPAPACSGIHRAICIGAGDLWILONHLLA 411
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 372 NKFDSDVPIILIKCLAEKLEEQ 393
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 NKFDPRVDNVLQCLEEYLRRK 433
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
035981 PRELIMINARY; PRT: 428 AA.

```

```

AC 035981;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GLUCOSAMINYL (N-ACETYL) TRANSFERASE 1, CORE 2 (BETA-1,6-N-
DE ACETYLGLUCOSAMINYLTRANSFERASE).
GN GCNT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=SUBMAXILLARY GLAND, AND KIDNEY;
RX MEDLINE=98001705; PubMed=9341170;
RA Sekine M., Nara K., Suzuki A.;
RT "Tissue-specific regulation of mouse core 2 beta-1,6-N-
RT acetylglucosaminyltransferase."
RL J. Biol. Chem. 272:27246-27252(1997).
DR EMBL: D87333; BAA22998.1;
DR MGD: D87332; BAA22998.1;
DR MGD: MGI:95676; Gcnt1.
DR InterPro: IPR003406;
DR Pfam: PF02485; Branch: 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 428 AA; 49831 MW; F0D10A6ADB23B92C CRC64;

Query Match 38.0%; Score 833; DB 11; Length 428;
Best Local Similarity 43.8%; Pred. No. 2.6e-57;
Matches 166; Conservative 66; Mismatches 127; Indels 20; Gaps 6;

Qy 27 DEYRYEVNCSGITDEPPEIGK-----SLERRDDIDLEDVYAMTSDCDIYQTLRG 80
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 DDPYSVNVNCTKILOGPDEIQRKLEILTVQFKRR--RWTPHDITNMRDCASTIRTK 108
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 81 YAKIVSKSEKSFPIAYSIVHVKDAIMVERLIHAIYGNHNYCIHYDKADTERVAMN 140
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 YIYEPPLKEVGRPIAYSTIVHKKIEMDLRLRAIYMPNFCINHYDKAESEFLAAVG 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 141 LAKCFNSNFIASKLEAVEYAHISRLQADLNCSDLKSSIQMKYVYNLCGDFPLKSNE 200
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 IASCFDNVFAVQSLSEVYVYASMSRVQADLNCMDLIRMANMKRYLLNCGMDFPIKTLE 228
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 201 LVSELKINGANMLTVPKPNKSLERFTYHHLRVPYEVYKLPRTNISKAPPHNIQI 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 IYRKIKCSTENLELETKMPNKEERMK-----KRYAVDGR-LTGTIVKAPPIKPL 282
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 261 FVGSATFVLSQAFVYIFNNSIVQDFPMSKDTYSPDEHFWATLIRVGPGEI-SRQAO 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 FSGSAFVVYREVYGVLENENIQKLEMAADTYSDEFLMATIQRIPEVPSFSPSNKY 342
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 320 DVSQDLSQSTRLYKMYIGSFF-----YPSCTGSHLSVCYIAGAEIRMLIKDGHMANKF 374
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 DLSDMNAIRFAVQWYHGGDVSMPAPACSGIHRAICIGAGDLWMLRQHLEFANKF 402
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 375 DSKVDPIILIKCLAEKLEEQ 393
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 DIDVDFPAIQLDEHLRHK 421
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
09HCVB PRELIMINARY; PRT: 314 AA.
AC 09HCVB;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE D1153D9.2 (A NOVEL PROTEIN SIMILAR TO BETA 1,6-N-
DE ACETYLGLUCOSAMINYLTRANSFERASE.) (FRAGMENT).
GN D1153D9.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

Db 127 QPONOFLAVDGNSSVEEIGLVRMLRCYGNIOYFIIDEIRMCCEYLTSV---FQCVDY 183
QY 175 LKSSIQKKYYINLCGDFPLKSNPELSEKTLGNAMLETYKPPNSKLERFYHHELR 234
Db 184 LAKLPDMKRYOYLSGVADPKSNLEMIKRLKALNG-----SFMET- 225
QY 235 RVPYEVKLPRTNISKAPPHNIOIFVGSAYFVLQAQVYIFNNSIVODEFAMSKDTY 294
Db 226 -LPFEFYRLNKRPMWSSPLPKTSL---SATF---SRKSANFVNSEKLEQIDFLRGTT 279
QY 295 SPDEFKATILRVPGI-----PGEISRSADQVSDL-QSKTLVYK-----NYE 337
Db 280 CADESLMKTIANPKYIKFSELPMPGPDARAWIKHNKRRTRGKLGXGENOKIDNGYV 339
QY 338 GFFYF-----SCTGSLRSYCIYGAELRWLIKDGHWFKNFKDSKYDPILIKLAETL 390
Db 340 SRQOYVNRAPYKCKGYRSLSCVGVYDLPNLIRHMLVAHKLYFSYQPAFACLVENS 399
QY 391 EQQ 393
Db 400 ROK 402

```

```

RESULT 9
002314 PRELIMINARY; PRT; 401 AA.
ID 002314
AC 002314
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE T15D6.2 PROTEIN.
GN T15D6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 283125; CAB05620.1;
DR InterPro; IPR003406;
DR Pfam; PF02485; Branch: 1.
SQ SEQUENCE 401 AA; 45608 MW; EID3F8AD84C0873D CRC64;

```

```

Query Match 13.5%; Score 286; DB 5; Length 401;
Best Local Similarity 23.4%; Pred. No. 2.7e-15;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;
QY 23 THKDEVREVNCSGIYEOEPLKSLRIRRDITLDDDDVAVMTSCDIYQTLRGYA 82
Db 34 TDLDDL--EINTNTL--QGLKNEKQELINTKMI---EDKLMTSTDCCHVXKSMFRN 86
QY 83 QKLVSEKESFPYLAIVLVYHKDAIMVERLIHALYNOHNITCIHYDKKADPTKRVAMNIA 142
Db 87 EYPLSEEARFPLSYGLVYKELSOVLFMLSIYQPNQYCIYAVGENSASTLLEELS 146

```

```

QY 143 KCFSNIFIAKLEAVEYAHISRLQADLNCISDLKSSIQKKYYINLCGDFPLKSNPEL 202
Db 147 DCFPN-----KRPIITWGSYEIINSYDDCKEFLSHLKSMMKRYQYLSGVDPILKTLNAV 201
QY 203 SELKTLGNAMLETYKPPNSKLERFYHHELRVPEYVYKLPRTNISKAPPHNIOIYV 262
Db 202 RLKSLNGTANVVEIKYENRLL-----LQNETESPFLK 237
QY 263 GSAFVPLQAQVYIFNNSIVODEFAMSKDTYSPDEFKATILRVPGI---IPGEISRSQA 319
Db 238 SSLSLIPKRAANYLASSIIPQDLLEFLRTWVADGFGNGLDVGSLN----- 293
QY 320 DVSDLSKTRLVKWNYYEGFPCSTGSLRSYCIYGAELRWLIKDGHWFKNFKDSKYD 379
Db 294 -----FEEHQIFESGCHNHMKDSCVFGIGDVSNTLQAKALVAHKLYLISE 340
QY 380 PILIKL 386
Db 341 PEAFVCL 347

```

```

RESULT 10
019729 PRELIMINARY; PRT; 454 AA.
ID 019729
AC 019729;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE F22D6.12 PROTEIN.
GN F22D6.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 271262; CA95816.1;
DR InterPro; IPR003406;
DR Pfam; PF02485; Branch: 1.
SQ SEQUENCE 454 AA; 52822 MW; FA0134DBF31D0925 CRC64;

```

```

Query Match 13.3%; Score 291; DB 5; Length 454;
Best Local Similarity 24.1%; Pred. No. 7.9e-15;
Matches 99; Conservative 73; Mismatches 174; Indels 64; Gaps 14;
QY 17 FVRNRYTHVDEV---RYVNCSGIYEOEPLKSLRIRRDITLDDDDVAVMTSDCQ 73
Db 40 FLKSICGYAQLKVEKESFPYLAIVLVYHKDA-----IMVERLIHAI 115
QY 74 IYQTLFRNTNPMASAEKHFPLSYGLVYKELSOVLFMLSIYQPNQYCIYAVGENSASTLLEELS 154
Db 95 TUKTLFRNTNPMASAEKHFPLSYGLVYKELSOVLFMLSIYQPNQYCIYAVGENSASTLLEELS 154

```

[illegible]

```

RESULT 11
Q020406 PRELIMINARY; PRT: 472 AA.
ID Q020406;
AC Q020406;
DT 01-NOV-1996 (TREMBLrel, 01, Created)
DT 01-NOV-1996 (TREMBLrel, 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel, 16, Last annotation update)
DE F44F4.6 PROTEIN.
GN F44F4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae.
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RA Coles L.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Boxfield J., Burton J., Connell M., Copestake T., Cooper J., Coulson A
RA Cresswell M., Dear S., Du Z., Dublin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Hilder M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shewkeen B.,
RA Smalton J., Smith A., Sonnhammer E., Staden A., Sulston J.,
RA Waterson A., Weisstock L., Wilkinson-Spoot J., Wohlman P.,
RA Watson A., Weisstock L., Wilkinson-Spoot J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans."
DR EMBL: Z37092; CAA8545.1; -
DR InterPro: IPR003406; -
DR Pfam: PF02485; Branch 1.
SQ SEQUENCE 472 AA; 55412 MW; 36A18D96D9C6245A CRC64;

```

```

Query Match 13.1% Score 288.5; DB 5; Length 472;
Local Similarity 22.0%; Pred. No. 1.3e+14;
Best Local 91; Conservative 78; Mismatches 175; Indels 69; Gaps 13
Matches

QY 40 EOEPEIGKSLR--TRRDI-----IDLEDDVYAMTSDCDIYQTLRGYAKLYSK 88
      |  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 63 EKIPEECNVTNGTEKRRKISRARQMDHFDWYEHIFSRNVCSTIDKYFNTRIPSSR 122
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 89 EEKSEPIASLYVHKDAIWERLIIHAIYNQHNYYCIHYDKAPDFKVAAMNLIACFSN- 147
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 123 EEAEPYPLAGLVVYKTIYQVLTQMSLFYQPMHFCITVDQSPNEKSVIQAIPSCFPNM 182
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 148 -FIASKLEAVEYAHISRLQADLNCISDILKSSIQMKYIYNLCQGDPLKSNFELVSEIK 206

```

Db 183 HAVIG-----EPSOMSGFGLIKLVYTCFPMMLKSKOKMKRYOYLSGDLITRINLEMRIFK 239

QY 207 KLGAMMETVAPFPMKIERFTYHHELRPVPEYVKLPITFNISKEAPHNIQIFVGSAY 266

Db 240 ALNGS-----AMTDVSTF-----EYDRY-----KMGESVLP--MPVYKSSMS 275

QY 267 FVLSQAFVKYIINNNSIVODFFAMSKDITYSPDEHFATLIRVG---IGELISKADQVSD 32

Db 276 VVPRREGADYLLISPRVOKLKLYSTKIWDIESHWSTVIGSPALLPVPST--RYRDLW 333

QY 324 LQSKTRL-----VKMYVEGF-----FYPSCTGSHLSRVCIYGAALMKLKDGH 365

Db 334 LRKNFKLRPEYFNATNSIGTSTIGRYQWGMOKCEGKYVDESCVGVDEIEITRPEL 393

QY 370 FAKKFPISKYDPLILKICLAEKLEEO-----RDWITLPEKIFMDRINT 412

Db 394 VAKKILIEQPAAPFMCMLKEVRRRSLSDPAHLPSARSYSQMPVELVYQKAT 446

RESULT	12			
0002315		PRELIMINARY:	PRF:	402 AA.
AC	0002315			
DT	01-JUL-1997 (TREMBLrel, 04, Created)			
DT	01-JUL-1997 (TREMBLrel, 04, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel, 16, Last annotation update)			
DE	tl5d6.3. PROTEIN.			
GN	tl5d6.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;			
OC	Rhabditidae; Pelodicerinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Dobson R;			
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEBLINE=94150718; PubMed=790638;			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Boxfield J., Burton J., Connell M., Copey T., Cooper T., Coulson A.			
RA	Croxton M., Dear S., Du Z., Dublin R., Favell A., Fulton L.,			
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,			
RA	Jones M., Keshaw J., Kirsten J., Laister N., Letellier P.,			
RA	Lighting J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,			
RA	Parsons L., Perry C., Riken L., Roopra A., Saunders D., Shownkeen F.			
RA	Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.			
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,			
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Woulman P.,			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans".			
RL	Nature 368:32-38 (1994).			
DR	EMBL: 283125; CAB05621.1; -			
DR	InterPro: IPR003406; -			
PI	Plan; PF02485; Branch 1.			
SEQ	SEQUENCE 402 AA; 46135 MW; F6AD721976703307 CRC64;			

Query Match	12.1%	Score 265	DB 5	Length 402
Best Local Similarity	24.7%	Pred. No. 7	3e-13	
Matches	93	Conservative	64	Mismatches 157, Indels 62, Gaps 13,
QY	23	THVKDEVREYVNCSTGYEOPLEIGKSLERDRDIDDEDDVAMTSCDIIQYLRGVA	82	
Db	44	TLDLDDL-QINCTNLQGFKE--NTLILITKML--ENKMNSTDCQTLTSEFRN	96	
QY	83	OKLVSKEKSPFLAYSLVYHKDAIMERLIIHAYNQHNYCIHYDKAPDTEKVAANNLTA	142	
Db	97	KVPLSEEARPFPLSFELLYVKELQYLFLLSSIQYQNEYCAVGENSAPAFILIKELA	156	
QY	143	KCSNFIKSLKEAVLYAHISRLQADLNLSDLKSIQWKYVYNIGQGFPLKSFELY	202	

```

Db 157 NCFPN-----KRPPIKMGSEFINSYGYCLEFELSHLKSMDKRYQYLSGVDIPILKTNLEW 211
QY 203 SELKINGAMLETVKPPNSKLEFTHHELRPRPYEVYKLPITNTSKAPPNIOIFV 262
Db 212 RILKRLNG-----TVNIGIS-----TYEDRL-----LNGKNTSP---LPLFK 247
QY 263 GSAFVLSQAFVKYITFNNSIVQDFEAWSKDYPSPDEHFMT-----LIRVPGIGEISR 316
Db 248 SLSLIPRKANVLSSSVPQQLLEFLRTWVADGEFGWGLFEGNKDLFVNGPSFNENDP 307
QY 317 SAQVSDLOSRTLYKWM-----NYGSEFPYSCGSHLSVCIYGAELRWLIKDGHW 370
Db 308 LTNMGNVYSRHOL--VWESECHNYMKD-----RS-CVFGIGDVPNLMKRSBALV 353
QY 371 ANKFDSKVDPIILKCL 386
Db 354 AHKLVESEPEAFCL 369

RESULT 13
C02309 PRELIMINARY; PRT; 459 AA.
ID 002309
AC 002309:
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE T09E11.9 PROTEIN.
GN T09E11.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCSL_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RL McClay K.;
RN Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Beers M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Smailson N., Percy C., Riffen L., Roopra A., Saunders D., Showkeen R.,
RA Spaldon J., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thelery-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sprat J., Wohlman P.,
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
RL Nature 368:32-38(1994).
DR EMBL: 281147; CAB03536.1;
DR InterPro: IPR003406;
DR Pfam: PF02485; Branch: 1.
SQ SEQUENCE 459 AA; 53476 MW; A514592B10F1A8E CAC64;

Query Match 11.2%; Score 245; DB 5; Length 459;
Best Local Similarity 22.8%; Pred. No. 3.2e-11;
Matches 106; Conservative 74; Mismatches 202; Indels 82; Gaps 17;

QY 5 YLVEVSLSTSPVRRKRYHVDVRYEYVNCGIY--EEPELIGKSLFIRROLDIED 62
Db 25 VYVIMES-SKSLGRN-----EQNSYALYNLRNENYOREATINELRDRFAKLES 72
QY 63 -----DVVAMTSDDDIYQTLGNGYQKLVSEKESFP----- 94
Db 73 KIEHFYSKTYRRSEFAHDCGRIISGDKDIQTYSG--ENRIPVWENNPNMCSAVMD 130
QY 95 -----TAYLVVHKDAIMYERLIIAYNOHNHYCHYDKKAPDTKVAMN 139
Db 131 RIIPHDIILPLKNGVAFARIYKDYELVEKOVMSYHPQNSCFADIKKAPTRKNMR 190

```

```

QY 140 NLAKFSNIFASKLEAVEVAHSRLQADLNCISDLKSSIQMKYVINLCGDFPKSNE 199
Db 191 AMACIPNVLLPDPQEPIDSNHNVLAHLNCLRALINRP--GMNVAAMLQNDLTKSY 249
QY 200 ELVSELKRLNGANMLETVPKPPNSKLEFTHHELRPRPYEVYKLPITRINISKEAPPHIQ 259
Db 250 ELQVYEWEMGANDVELL--PEAQRLDEENFKMD---PSLMKFPDESKVDETIINELIK 304
QY 260 IFVGSAYFVLSQAFVKYITFNNSIVQDF--AWSQDYSPEHFMT--LIRVPGIGEISR 316
Db 305 FSKGIVQGSKRRAVWMTKRVNLSTYIDQWNGKGVDEMLSSIQIAFLMGH-- 362
QY 317 SAQVSDLOSRTLYKWMYEGFYPSCGSHLR--SVCITYGAELRWLIKDGHWANKFD 375
Db 363 TDQCLKEGKEPFPMYQWNEDESY---CASKYKRVHVCILGIELRSVASFPLTMNKML 419
QY 376 SKVDPIILKCLAELEO---ORDWITLPSKELEMDRLTTTS 415
Db 420 PPFDSITTECTAELLYNRTFMQNDH---PLEEYI-KMWTYS 459

RESULT 14
C09M034 PRELIMINARY; PRT; 876 AA.
ID 09M034
AC 09M034; C09M035; Q24116;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG17771 PROTEIN.
GN CG17771 OR CG17772 OR ANON-12.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RA TISSUE=IMAGINAL DISK;
RA Bessarab D.A.; Sun H.Y.;
RN Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
[1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RA STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.G., Kottman J.R., Yandell M.P., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Adiri J.F., Aghayani A., An H.-J., Andrews-Plannkoc C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.S.M.,
RA de Pablo B., Delcher A., Deng Z., May S.A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
RA Jaitani M., Kalush F., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Klumel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Mostnefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzay D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```



Fri Aug 3 09:38:18 2001

us-09-645-192-2\_copy\_39\_453.rpt

---

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compaq Inc.

OM protein - protein search, using SW model

Run on: August 2, 2001, 14:50:18 ; Search time 16.71 Seconds

(without alignments)  
850.750 Million cell updates/sec

Title: US-09-645-192-2\_COPY\_39\_453  
Perfect score: 2194  
Sequence: 1 FOKDIYEVSLSTSPFYRN.....DWITLPSEKLFMDNLTTS 415

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	849.5	38.7	428	1 G6NT_HUMAN	002742 homo sapien
2	837	38.1	428	1 G6NT_MOUSE	009374 mus musculu
3	831.5	37.9	427	1 G6NT_BOVIN	092180 bos taurus
4	730.5	33.3	400	1 BGIB_MOUSE	097402 mus musculu
5	711	32.4	400	1 BGIB_HUMAN	006430 homo sapien
6	711	5.3	895	1 YA54_METJA	058454 methanococ
7	108.5	4.9	661	1 WHI3_YEAST	034781 saccharomyc
8	102.5	4.7	654	1 BFR2_HUMAN	001742 homo sapien
9	102.5	4.7	821	1 FGR2_HUMAN	021802 homo sapien
10	102.5	4.7	821	1 FGR2_MOUSE	021803 mus musculu
11	100.5	4.6	2607	1 BACB_BACLI	068007 bacillus 11
12	99.5	4.5	666	1 VOO1_VACCC	021093 vaccinia vi
13	99.5	4.5	823	1 CEK3_CHICK	018491 gallus gall
14	97.5	4.4	1314	1 TETX_CLOTE	04958 clostridium
15	97.5	4.4	2054	1 YCF2_PINTH	041653 pinus thub
16	97	4.4	440	1 GAAP_HUMAN	000531 homo sapien
17	96	4.4	1036	1 YOB6_CAREL	009292 caenorhabdi
18	95	4.3	701	1 SYGB_HELPJ	092489 heliobacte
19	94	4.3	1344	1 IF3A_MOUSE	023116 mus musculu
20	94	4.3	1675	1 CLH1_HUMAN	000610 homo sapien
21	94	4.3	1675	1 CLH_BOVIN	049591 bos taurus
22	94	4.3	1675	1 CLH_RAT	041442 rattus norv
23	93.5	4.3	503	1 SECA_PORPU	051381 porphyra pu
24	93	4.2	503	1 CPV1_SHEEP	092488 ovis aries
25	93	4.2	520	1 Y208_METJA	060270 methanococ
26	93	4.2	576	1 POLO_DROME	052304 drosophila
27	92	4.2	908	1 DPO1_BORBU	051498 borrelia bu
28	92	4.2	1103	1 CYGF_BOVIN	002740 bos taurus
29	92	4.2	1382	1 IF3A_HUMAN	041452 homo sapien
30	91.5	4.2	720	1 SPOT_MYCGE	047550 mycoplasma
31	91.5	4.2	1956	1 ATX1_PLAFA	004956 plasmodium
32	91	4.1	364	1 D3B_MYCGE	047247 mycoplasma
33	91	4.1	1037	1 Y0J8_YEAST	012456 saccharomyc

34	91	4.1	1077	1 DN2L_HUMAN	P51530 homo sapien
35	90.5	4.1	560	1 INR1_SHEEP	Q28589 ovis aries
36	90.5	4.1	759	1 TEAL_YEAST	P47988 saccharomyc
37	90.5	4.1	850	1 CYAA_YERPE	P40127 yeastinia pe
38	90.5	4.1	936	1 MSH4_HUMAN	015457 homo sapien
39	90.5	4.1	1036	1 Y414_MYCGE	P47653 mycoplasma
40	90.5	4.1	2280	1 YCF2_TORAC	P09976 nicotiana t
41	90	4.1	415	1 EPIG_YEAST	P29547 saccharomyc
42	90	4.1	503	1 CPV2_PIG	P79430 sus scrofa
43	90	4.1	605	1 GLMS_THEMA	Q57425 t glucosami
44	90	4.1	1324	1 VGL2_CVMA5	P11224 murine coro
45	90	4.1	1655	1 N188_YEAST	P52593 saccharomyc

## ALIGNMENTS

RESULT ID	1 G6NT_HUMAN	STANDARD:	PRT:	428 AA.
AC	002742:			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)			
DE	(CORE2-GLCNAC-TRANSFERASE) (CGNT) (CORE 2 GNT).			
GN	GCNT1			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-93028457; PubMed-1329093;			
RA	Bierhuizen M.F.A., Fukuda M.,			
RT	"Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6-GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen."			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Placenta;			
RX	MEDLINE-96078409; PubMed-7579796;			
RA	Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.,			
RT	"Genomic organization of core 2 and 1 branching beta-1,6-N-acetylglucosaminyltransferases. Implication for evolution of the beta-1,6-N-acetylglucosaminyltransferase gene family."			
RL	Glycobiology 5:417-425(1995).			
CC	- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.			
CC	- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + BETA-D-GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-GLUCOSAMINYL-1,6-N-ACETYL-D-GALACTOSAMINYL-R.			
CC	- PATHWAY: GLYCOSYLATION.			
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.			
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES AND MYELOID CELLS.			
CC	- SIMILARITY: TO T-BRANCHING ENZYME (IGNT).			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL: M97347; AAA55919.1;			
DR	EMBL: L41415; AAA96661.1;			
DR	PIR: A46293; A46293.			
DR	MTM: 600391;			
KW	Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;			

KW Golgi stack; glycoprotein.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 32  
 FT  
 FT  
 FT DOMAIN 33 428  
 FT CARBOHYD 52 52  
 FT CARBOHYD 58 58  
 FT CARBOHYD 95 95  
 SQ SEQUENCE 428 AA; 49784 MW; 7C9CEB51DA0A60 CRC64;

Query Match 38.7%; Score 849.5; DB 1; Length 428;  
 Best Local Similarity 43.0%; Pred. No. 42e-55;  
 Matches 181; Conservative 70; Mismatches 129; Indels 41; Gaps 11;

1 POKDIYV-EYSLSTSPFVNRRTYHVKD---VRY-----EVCNCGIYEEPLEI 46  
 14 PTXYFVYVLTSLTFSVLR---IHQKPEFVSVRHLELAGENPSSDINCKYLGQDVNEI 70  
 DB  
 QY 47 GR-----SLEIRRDIILEDVVAMTSDCIYOTLGYAKIYKREKSPFIATSLY 100  
 DB  
 DB 71 QKVKLELTVYKFKRP--RWTPTDYIMNTSDCSSFKRRKIVYELSKAEAPFIATSLY 128  
 QY 101 VHKDAIYVERLIIHAYNOHNYICIHDRKADTEFKVAMNNLAKPSNIFIAKLEAVEYA 160  
 DB 129 VHKIEMLDRLIRAIYQNFYCVHVDTSKESDYLAAMGJASCFSNVFAVSLAESVYA 188  
 QY 161 HISLQADNCLSDLKSIQKRYVINLGGDFPLKSNELVELSKKNGAMLETKRP 220  
 DB 189 SMSVQADLCKMDKDLVAMSNMRYLNLGMDPIRTINLEIYKLLKNGENLETERHP 248  
 QY 221 NSKLEPTVHEHLEIRVPEYEV--KLPIRTNISKEAPPHNIQIFVSGAYVLSQAFVKYI 278  
 DB 249 SHKEEM-----KKRIYEVNCKL-TNIGVAMLPLEPLFSGSAYFVSEYGYVL 300  
 QY 279 NNSIVQDFPAMSKDTYSPDEHFNATLIRVGIGETISRAQ--CVSDLOKSTRLVKNAYE 337  
 DB 301 QNEKIQIKLMAQDTYSPDEYLAFTIQRIPVGSPLASHKYDLSDMAVAFVKKOYEE 360  
 QY 338 GFF-----YPSCTGSHLRVSCIYGAELRWLIDGHWFAKRPDSKYDPIILKLAKEE 392  
 DB 361 GDVSKGAPYPCGSGVHYRASCIFGAGDLNMLRKHLFANKPFDVVDLFIQCLDELHRR 420  
 QY 393 Q 393  
 DB 421 K 421

RESULT 2  
 G6NT\_MOUSE STANDARD; PRT; 428 AA.  
 ID G6NT\_MOUSE  
 AC Q09324;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-  
 DE ACETYLGUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)  
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT).  
 GN G6NT1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=DA/2;  
 RA Warren C.E., Smockler D.S., Dennis J.W.;  
 Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-  
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-  
 CC GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-  
 CC GALACTOSAMINYL-R.

CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGBT).  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: 019265; AAA60948.1; -  
 DR MGI:95676; G6nt1.  
 DR Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; Glycoprotein.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 32  
 FT  
 FT  
 FT DOMAIN 33 428  
 FT CARBOHYD 58 58  
 FT CARBOHYD 95 95  
 SQ SEQUENCE 428 AA; 49849 MW; F3FED2517B27AA61 CRC64;

Query Match 38.1%; Score 837; DB 1; Length 428;  
 Best Local Similarity 43.8%; Pred. No. 3.5e-54;  
 Matches 166; Conservative 67; Mismatches 126; Indels 20; Gaps 6;

27 DEVREVCNCGIYEEPLEIQQ-----SLEIRRDIILEDVVAMTSDCIYQTLRG 80  
 DB 51 DDPYVNVCTKLQDDPEIQVKLELTVQKRP--RWTPTDYIMNTPCASFIRTK 108  
 QY 81 YAKIYKREKSPFIATSLYVHKDAIYVERLIIHAYNOHNYICIHDRKADTEFKVAMNN 140  
 DB 109 YVEPLTEKEVGFPIATSLYVHKIEMLDRLIRAIYQNFYCVHVDTSKESDYLAAMG 168  
 QY 141 LAKCSNIFIAKLEAVEYAHISRLQADNCLSDLKSIQKRYVINLGGDFPLKSNFE 200  
 DB 169 IASCFDNVFAQLSSVYASWTYKADNCKMDLYRMANMKYILNLGMDPIRTINLE 228  
 QY 201 LVSELKKNLGMAMLETVKPPNSKLERTYHLEIRVPEYEVYKLPRTNISKEAPPHNIQI 260  
 DB 229 IYRKILKCSGENNLETERKPPNKERRMK-----KRYAVVDGKL-TNIGIVAKPPLKTP 282  
 QY 261 EVGSATFVLSQAFVKYIIFNNSIVQDFPAMSKDTYSPDEHFNATLIRVGIGETISRAQ 319  
 DB 283 FGSATFVTVTRYGVGLTENENIQKLEMAQDTYSPDEYLAFTIQRIPVGSPLSSNKY 342  
 QY 320 DVSDDQSTRLVKNAYEEGF-----YPSCTGSHLRVSCIYGAELRWLIDGHWFAK 374  
 DB 343 DLSDMAIARFVKQWFEBEDVSGNAPYPCGSGVHYRASCIFGAGDLNMLRHLFANKF 402  
 QY 375 DSKVDPILKICAEKEEQ 393  
 DB 403 DMVDVDFEALQCLDELHRRK 421

RESULT 3  
 G6NT\_BOVIN STANDARD; PRT; 427 AA.  
 ID G6NT\_BOVIN  
 AC Q92180;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-  
 DE ACETYLGUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)  
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).  
 GN G6NT1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Li C.M., Cheng P.W.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-  
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-  
 CC GALACTOSYL-1,3-(N-ACETYL-D-GALACTOSAMINYL-1,6)-N-ACETYL-D-  
 CC GALACTOSAMINYL-R.  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U41320, AA83244.1;  
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; Glycoprotein.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 32  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 58 58  
 FT CARBOHYD 58 58  
 FT CARBOHYD 95 95  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 427 AA; 49733 MW; 41B9CBF948D4196 CRC64;  
 Query Match 37.9%; Score 831.5; DB 1; Length 427;  
 Best Local Similarity 45.6%; Pred. No. 8,8e-54;  
 Matches 170; Conservative 64; Mismatches 118; Indels 21; Gaps 7;  
 OY 33 VNCSGIYEOEPLLEIGK---SLEIRRDITIDEDDDVAVMTCSDIYQTLRGYAKLYSK 88  
 DB 57 INCKTLQGDVDEIOKVKLESLTVFKKRAWTNDYDINMTCDCASFTRKRYITEPLSK 116  
 OY 89 EKSPPIASLVVHKDAIYVERLIIAHYNQHYCIHYDRKAPDFFKAMNNLAKCFENI 148  
 DB 117 EEEGPIAISIVYHKKIEMDLRLAIIYWPQFYCIHWDAKSEKSFLLAAVAGISCFENV 176  
 OY 149 FASLKEAVEVAHISRLQADNCLSDLKSSIQWKYVINLGGDFPLKSNFELVSELK 208  
 DB 177 FVASQLESIVYASWSRVQADLNCMODLYOMNAGMYLLNLCMDPFIKXINLEIVKRL 236  
 OY 209 NGANMETVKKPNSKLERTYHHELRVPEYVKLPPIRTNI--SKAPPHNIOIFVSAY 266  
 DB 237 MGENNLETAKMPSHKERKKH-----YEVYNGKL-TNMGTDKIHPLEPLISGSNH 288  
 OY 267 FVLSQAFVYIYFNNSIVQDFEAFMSKDYSPDEHFNATLLRVGIPGELISRSAQ-DVSDIQ 325  
 DB 289 FVVSREYVEVLQNONIQOFEMAKDTPSPDELMATIORIEVEVPSLSLYKYPTSDMQ 348  
 OY 326 SKRLVKNMYIEGFF-----YPSCTGSHLRVSCIYGAALRLKINGCHFAKPKPSKVP 380  
 DB 349 AIARFVKWQYFEGDVSKGAPPPCS-VHVRSYCVFSGADLMLHVAHFLPAKFDID 407  
 OY 381 ILIKCLAELKEEQ 393  
 DB 408 FAIQGCDHEHLRHK 420  
 RESULT 4  
 BGIB\_MOUSE STANDARD; PRT; 400 AA.  
 ID BGIB\_MOUSE  
 AC P97402;

DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE N-ACETYLGLUCOSAMINYL BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE  
 DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)  
 DE (IGNT) (LARGE I ANTIGEN-FORMING BETA-1,6-N-  
 DE ACETYLGLUCOSAMINYLTRANSFERASE).  
 GN GCNT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97280061; PubMed=9134435;  
 RA Magnat A.D., Fukuda M.;  
 RT Expression of the large I antigen forming beta-1,6-N-  
 RT acetylglucosaminyltransferase in various tissues of adult mice.";  
 RL Glycobiology 7:285-295(1997).  
 CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED  
 CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I  
 CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED  
 CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-  
 CC -1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-D-  
 CC GLUCOSAMINYL-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U68182; AA839621.1;  
 DR MGD: MG11100870; Gcnt2.  
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; Glycoprotein.  
 FT DOMAIN 1 6  
 FT TRANSMEM 7 25  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 37 37  
 FT CARBOHYD 255 255  
 FT CARBOHYD 314 314  
 FT CARBOHYD 388 388  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 400 AA; 45507 MW; 193FE3862E3B9691 CRC64;  
 Query Match 33.3%; Score 730.5; DB 1; Length 400;  
 Best Local Similarity 40.0%; Pred. No. 2e-46;  
 Matches 156; Conservative 51; Mismatches 160; Indels 23; Gaps 6;  
 OY 5 IYLVESISTSPFYRNRYTHKDEYRVNCSGIYEOEPLIGSLERRDIDLEDD 64  
 DB 18 VCVASYVSLFPGDOSYOKLNSDSVRSLOYCSSFID-----GKSRFLWRKRL----- 64  
 OY 65 VVAMTSDCDIYQTLRGYAKLYSKSEKSFPIAYSLVYVHKDAIYVERLIIAHYNQHYCI 124  
 DB 65 MIHEKPSCTEYVTOGSHYITLPSQEDGFLAYMVVHHNDTFARLFRAIFMQNTYCV 124  
 OY 125 HYDRRAPDTEVYVAMNNLAKCFNSIFIASLFAVEAHISRLQADNCLSDLKSSIQWKY 184  
 DB 125 HYDEKATAIEFGAVQALVSCPPNFFLASKMPEVYVGGISRLQADNCLSDLKSSIQWKY 184  
 OY 165 VINLCGDFPLKSNFELVSELKINGANMLETVRPNKSLERTY-HHELRVPEYEVK 243  
 DB 165 AINTCGDFPLKTKNKELIYQIKGLKGNLTPGVLPRAHAIQRTYVHREHLSKELSY- 242







"K-sam, an amplified gene in stomach cancer, is a member of the heparin-binding growth factor receptor genes.";  
Proc. Natl. Acad. Sci. U.S.A. 87:5983-5987(1990).  
[5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92212948; PubMed=1313574;  
RA Katon M., Hattori Y., Sasaki H., Tanaka M., Sugano K., Yazaki Y.,  
RA Sugimura T., Terada M.;  
RT "K-sam gene encodes secreted as well as transmembrane receptor  
RT tyrosine kinase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:2960-2964(1992).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=93016048; PubMed=1400433;  
RA Dell K.R., Williams L.T.;  
RT "A novel form of fibroblast growth factor receptor 2. Alternative  
RT splicing of the third immunoglobulin-like domain confers ligand  
RT binding specificity.";  
RL J. Biol. Chem. 267:21225-21229(1992).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RA Steinberger D., Mueller U.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP VARIANTS CS HIS-340; ARG-342; SER-342; TYR-342 AND CYS-354.  
RX MEDLINE=95078932; PubMed=7987400;  
RA Reardon W., Winter R.M., Rutland P., Pulleyn L.J., Jones B.M.,  
RA Malcolm S.;  
RT "Mutations in the fibroblast growth factor receptor 2 gene cause  
RT Cronzon syndrome.";  
RL Nat. Genet. 8:98-103(1994).  
RN [9]  
RP VARIANTS CS CYS-328 AND CYS-347. AND VARIANT JMS GLY-344.  
RX MEDLINE=95179174; PubMed=78741170;  
RA Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J.,  
RA Charnas L.R., Jackson C.E., Jaye M.;  
RT "Jackson-Weiss and Cronzon syndromes are allelic with mutations in  
RT fibroblast growth factor receptor 2.";  
RL Nat. Genet. 8:275-279(1994).  
RN [10]  
RP VARIANTS CS  
RX Oldridge M., Wilkie A.O.M., Slaney S.F., Poole M.D., Pulleyn L.J.,  
RA Rutland P., Hockley A.D., Wake M.J.C., Goldin J.H., Winter R.M.,  
RA Reardon W., Malcolm S.;  
RT "Mutations in the third immunoglobulin domain of the fibroblast growth  
RT factor receptor-2 gene in Cronzon syndrome.";  
RL Hum. Mol. Genet. 4:1077-1082(1995).  
RN [11]  
RP VARIANT PS ALA-321.  
RX MEDLINE=95235551; PubMed=7719333;  
RA Lajeunie E., Wei M.H., Bonaventure J., Munnich A., le Merrer M.,  
RA Renier D.;  
RT "FGFR2 mutations in Pfeiffer syndrome.";  
RL Nat. Genet. 9:108-108(1995).  
RN [12]  
RP VARIANTS AS TRP-252 AND ARG-253.  
RX MEDLINE=95235562; PubMed=7719344;  
RA Wilkie A.O.M., Slaney S.F., Oldridge M., Poole M.D., Ashworth G.J.,  
RA Hockley A.D., Hayward R.D., David D.J., Pulleyn L.J., Rutland P.,  
RA Malcolm S., Winter R.M., Reardon W.;  
RT "Apert syndrome results from localized mutations of FGFR2 and is  
RT allelic with Cronzon syndrome.";  
RL Nat. Genet. 9:165-172(1995).  
RN [13]  
RP VARIANTS PS PRO-341; ARG-342 AND TYR-342.  
RX MEDLINE=95235563; PubMed=7719345;  
RA Rutland P., Pulleyn L.J., Reardon W., Baraister M., Hayward R.,  
RA Jones B., Malcolm S., Winter R.M., Oldridge M., Slaney S.F.,  
RA Poole M.D., Wilkie A.O.M.;  
RT "Identical mutations in the FGFR2 gene cause both Pfeiffer and  
RT Cronzon syndrome phenotypes.";  
RL Nat. Genet. 9:173-176(1995).  
RN [14]  
RP VARIANTS CS/JMS/PS.  
RX MEDLINE=96203627; PubMed=8644708;  
RA Meyers G.A., Day D., Goldberg R., Daentl D.L., Przylepka K.A.,  
RA Abrams L.J., Graham J.M. Jr., Feingold M., Moeschler J.B.,  
RA Ramsley E., Scott A.F., Jabs E.W.;  
RT "FGFR2 exon IIIa and IIIc mutations in Cronzon, Jackson-Weiss, and  
RT Pfeiffer syndromes: evidence for missense changes, insertions, and a  
RT deletion due to alternative RNA splicing.";  
RL Am. J. Hum. Genet. 58:491-498(1996).  
RN [15]  
RP VARIANTS CS CYS-105; GLU-338; CYS-351 AND ARG-384.  
RX MEDLINE=97101656; PubMed=8946174;  
RA Pulleyn L.J., Reardon W., Wilkes D., Rutland P., Jones B.M.,  
RA Hayward R., Hall C.M., Brunton L., Chun N., Lammer E., Malcolm S.,  
RA Winter R.M.;  
RT "Spectrum of craniosynostosis phenotypes associated with novel  
RT mutations at the fibroblast growth factor receptor 2 locus.";  
RL Eur. J. Hum. Genet. 4:283-291(1996).  
RN [16]  
RP VARIANTS CS ILE-331; ASP-336--ALA-337 DUPL AND TRP-356--THR-358 DEL.  
RX MEDLINE=97114301; PubMed=8956050;  
RA Steinberger D., Mulliken J.B., Mueller U.;  
RT "Cronzon syndrome: previously unrecognized deletion, duplication, and  
RT point mutation within FGFR2 gene.";  
RL Hum. Mutat. 8:386-390(1996).  
RN [17]  
RP VARIANT PS CYS-290.  
RX MEDLINE=97295073; PubMed=9150725;  
RA Tartaglia M., Valeri S., Velardi F., di Rocco C., Battaglia P.A.;  
RT "Trp290Cys mutation in exon IIIa of the fibroblast growth factor  
RT receptor 2 (FGFR2) gene is associated with Pfeiffer syndrome.";  
RL Hum. Genet. 99:602-606(1997).  
RN [18]  
RP VARIANT CS I-252. VARIANT AS F-252, AND VARIANT PS F-252--S-253.  
RX MEDLINE=97156222; PubMed=9002682;  
RA Oldridge M., Lunt P.W., Zackai E.H., McDonald-Mcginn D.M., Muenke M.,  
RA Moloney D.M., Twigg S.R.F., Heath J.K., Howard T.D., Hoganson G.,  
RA Gagnon D.M., Jabs E.W., Wilkie A.O.M.;  
RT "Genotype-phenotype correlation for nucleotide substitutions in the  
RT Ig1-IgIII linker of FGFR2.";  
RL Hum. Mol. Genet. 6:137-143(1997).  
RN [19]  
RP VARIANTS CS GLU-292.  
RX MEDLINE=97297373; PubMed=9152842;  
RA Steinberger D., Collmann H., Schmalenberger B., Mueller U.;  
RT "A novel mutation (A886G) in exon 5 of FGFR2 in members of a family  
RT with Cronzon phenotype and plagiocephaly.";  
RL J. Med. Genet. 34:420-422(1997).  
RN [20]  
RP VARIANTS CS VAL-276; CYS-301 AND SER-314.  
RX MEDLINE=98180879; PubMed=9521581;  
RA Steinberger D., Vriend G., Mulliken J.B., Mueller U.;  
RT "The mutations in FGFR2-associated craniosynostoses are clustered in  
RT five structural elements of immunoglobulin-like domain III of the  
RT receptor.";  
RL Hum. Genet. 102:145-150(1998).  
RN [21]  
RP VARIANTS AS TRP-252 AND ARG-253.  
RX MEDLINE=98112406; PubMed=9452027;  
RA Tsai F.-J., Hsu W.-L., Lin S.-P., Chang J.-G., Wang T.-R., Tsai C.-H.;  
RT "Two common mutations 934C to G and 937C to G of fibroblast growth  
RT factor receptor 2 (FGFR2) gene in Chinese patients with Apert  
RT syndrome.";  
RL Hum. Mutat. Suppl. 1:S18-S19(1998).  
RN [22]  
RP VARIANT PS CYS-351.  
RX MEDLINE=98358420; PubMed=9693549;  
RA Mathijssen I.M., Vaandrager J.M., Hoogeboom A.J.,  
RA Hesselink-Janssen A.L., Van den Ouweland A.M.W.;  
RT "Pfeiffer's syndrome resulting from an S351C mutation in the







```

Db      834 YKVIADQYGRPOV-----GVPGLLIGEGVARGYLNHET-LTK-----AAFYVD 879
QY      344 CTGSHLRVCICGAEL-RWLKIDGHWPFANKPDSKYDPLLKICLAEKLEOORDWITLPS 402
      880 ESEGER-----VYRTGDLARWLSDGNIEFLGRIDSQV---IGGYIELEIEE-----923
QY      403 EKLFMDRNL 411
      924 HRLMNDNI 932

RESULT 12
V001_VACCC STANDARD; PRT; 666 AA.
AC P21093;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE PROTEIN O1.
GN O1L.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_Taxid=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91021027; PubMed-2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RL Virology 179:517-563(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M35027; AAA48053.1; -
DR PIR; D42510; D42510.
SQ SEQUENCE 666 AA; 77577 MW; 46DF3299DDEB66EF CRC64;

```

```

Query Match 4.5%; Score 99.5; DB 1; Length 666;
Best Local Similarity 22.0%; Pred. No. 5.7; 11; Indels 105; Gaps 18;
Matches 75; Conservative 50; Mismatches 11;

```

```

QY      8 VEVLSTSPFRNRYTHVKEVREVCNCSGYEEOPELICKSLER-----54
      292 VEVNIISTYDF---RYROFADEFRIYIM---IKERQITMOSGDIRFRFRMSLRSTIIR 345
      55 RDIDLED-----DDVY-----AMSDCDIOTLRGTAQ 83
      346 KDTLSLEDILAHIDNARKNSKVSIEDVERLITSSFRNLPCVVRRTMLSDIDIKTKI--MVL 403
QY      84 KLVKREKESFPIAYSLVHNKAIWVERLHAIVQNHNYICGHYDRK-----ADTFKVA 137
      404 KIV-KDMKSCALITLSAI---KGIVTDTINVLK-----IHHNRANFKYLTSEKNEIA 455
Db      456 VCNCSRCL-----LEFRELKSVRCDLRTDDGLDLRLXLDLTRYALHGKINQMLIG 505
QY      138 MNNAKCFNSNFIASKLEAVEYAHISRLQADLNC-----LSDLKSSIQWKYVINLQCG 190
      456 VCNCSRCL-----LEFRELKSVRCDLRTDDGLDLRLXLDLTRYALHGKINQMLIG 505
QY      191 QDF--ELKSNFELVSELKKGANMLETVKRPNSKLEFRYTHHLLRVRPEYVYKLPRTN 248
      506 QRCWGPFLTEMLFENKRRKLN--NLMETIKIS-----DMLVYGHSI-----EKTILPIITDS 554

```

```

QY      249 ISKEAPPNHNIQIFVGSAYEVLSCAFYK-YIFNNSIYODEFA 288
      555 LSEKLSYDTMS-----VLNDQYAKIVIFNFTIIEYIIA 587
Db      555 LSEKLSYDTMS-----VLNDQYAKIVIFNFTIIEYIIA 587

RESULT 13
CEK3_CHICK STANDARD; PRT; 823 AA.
ID CEK3_CHICK
AC P18461;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TYROSINE KINASE RECEPTOR CEK3 PRECURSOR (EC 2.7.1.112).
GN CEK3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90332672; PubMed-2165604;
RA Pasquale E.B.;
RT "A distinctive family of embryonic protein-tyrosine kinase
RT receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:5812-5816(1990).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M35196; AAA48655.1; -
DR PIR; B35963; B35963.
DR HSSP; P11362; IEG1.
DR InterPro; IPR000719; -
DR InterPro; IPR001245; -
DR InterPro; IPR003006; -
DR Pfam; PF00047; Ig; 3.
DR Pfam; PF00069; PKINASE; 1.
DR PRINTS; PRO0109; TYRKINASE.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM_1.
KW Receptor; Glycoprotein; Tyrosine protein kinase; ATP-binding;
KW Transferrase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1..23
FT CHAIN 24..823
FT DOMAIN 24..379
FT TRANSMEK 380..400
FT DOMAIN 401..823
FT DOMAIN 58..117
FT DOMAIN 174..240
FT DOMAIN 273..351
FT DOMAIN 135..145
FT DOMAIN 483..772
FT NP_BIND 489..497
FT BINDING 519..519
FT ACT_SITE 628..628
FT MOD_RES 659..659
FT DISULFID 65..110
FT DISULFID 181..233
FT DISULFID 280..344
FT CARBOHYD 86..86

```

```

FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 823 AA: 92299 MW: 42833CC4EA02ED43 CRC64;

Query Match 4.5%; Score 99.5; DB 1; Length 823;
Best Local Similarity 24.0%; Pred. No. 7.5; Indels 103; Gaps 18;
Matches 78; Conservative 34; Mismatches 110;

QY 144 CFSNFIASKL-----EAVEAHISRLQADL--NCISDL-----IKSSIQWRYVIN 187
DB 493 CFGVVAEAVGIDKORPREAVTVA--VKMLKDATEKDLSDVSEMEMMKHKHNIIN 551
QY 188 L---CGDDFLKSNFELVSELKLNKNGMLETYKPPSKLERTYHHELRRVPEYKLP 244
DB 552 LLAGCTDGDPL--YVIVYASKNREYLRARRPPEME-----TSFDIKRVEEDQTF- 602
QY 245 IRTNISKAPPNNIOIFVGSAYEV-----LSQAFVKYIFNNSIYQDFEAWSKDTYSP 296
DB 603 -----KDIVSCTQLARGHEVLASCKCIHDLAARVLTENNVMKIADFGLARQIINI 656
QY 297 DEHFMAFLIRPGIIGETISSAQDVSLQSKTRYKKNYEGEFYPCQSHLRVCYIG 356
DB 657 DYKKTNGRLP-----VKWMADEAFEDRYVT--HOSDWSFG 692
QY 357 -----AAELRWLIKRGHWFAKPKSKYDPILIKCLAEKLEQORD-W 397
DB 693 VLMEITLGLSPYPCIPVELLEFKLKEGH---RMDKPN-----CTNE-LVMWMDCW 742
QY 398 ITLPSEK-LF-----MDRNTTTS 415
DB 743 QAVPSORPTEFKQVEDLDRLITLT 767

RESULT 14
TEXT_CLOTE STANDARD: PRT; 1314 AA.
ID TEXT_CLOTE STANDARD: PRT; 1314 AA.
AC P04958;
DR 13-AUG-1987 (Rel. 05, Created)
DR 13-AUG-1987 (Rel. 05, Last sequence update)
DR 15-JUL-1999 (Rel. 38, Last annotation update)
DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
OS Clostridium tetani.
OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87053814; PubMed=3536478;
RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA Weller U., Hudel M., Habermann E., Nlemann H.,
RT "Tetanus toxin: primary structure, expression in E. coli, and
RT homology with botulinum toxins."
RL EMBO J. 5:2495-2502(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-CN3911;
RX MEDLINE=87040747; PubMed=3774547;
RA Fairweather N.F., Lyness V.A.;
RT "The complete nucleotide sequence of tetanus toxin."
RT Nucleic Acids Res. 14:7809-7812(1986).
RN [3]
RP SEQUENCE OF 742-1314 FROM N.A.
RX MEDLINE=86085672; PubMed=3510187;
RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT "Cloning, nucleotide sequencing, and expression of tetanus toxin

```

```

RT fragment C in Escherichia coli."
RL J. Bacteriol. 165:21-27(1986).
RN [4]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=90201034; PubMed=2108021;
RA Kriegstein K., Henschen A., Weller U., Habermann E.,
RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT in tetanus toxin."
RL Eur. J. Biochem. 186:39-45(1990).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE=92037649; PubMed=1935979;
RA Kriegstein K.G., Henschen A.H., Weller U., Habermann E.;
RT "Limited proteolysis of tetanus toxin. Relation to activity and
RT identification of cleavage sites."
RL Eur. J. Biochem. 202:41-51(1991).
RN [6]
RP IDENTIFICATION OF ZINC-PROTEASE.
RX MEDLINE=93010948; PubMed=1396558;
RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA Montecucco C.;
RT "Tetanus toxin is a zinc protease and its inhibition of
RT neurotransmitter release and protease activity depend on zinc."
RL EMBO J. 11:3577-3583(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.F.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin."
RL Nature 359:832-835(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE=97475217; PubMed=9334741;
RA Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA Sax M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin."
RL Nat. Struct. Biol. 4:788-792(1997).
CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-1-PHE-77
CC BOND OF SYNAPTOBREVIN-2.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-1-PHE-77 BOND IN
CC SYNAPTOBREVIN.
CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC GANGLIOSIDE RECEPTORS.
CC -1- SIMILARITY: BELONGS TO THE TETANUS/BOTULINUM NEUROTOXIN
CC METALLOPROTEASES; ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04436; CAA28033.1; -
CC EMBL: M12739; AAA23282.1; -
CC EMBL: X06214; CAA29564.1; -
CC PIR: A25689; BTCLIN
CC PDB: 1AF9; 29-APR-98.
CC PDB: 1A8D; 14-OCT-98.
CC MEROPS: M27.001; -

```

```

DR InterPro: IPR000130;
DR InterPro: IPR000395;
DR Pfam: PF01742; Peptidase-M27; 1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydroxylase; Metalloprotease; Zinc; Plasmid;
DR 3D-structure.
FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
FT CHAIN 2 457 TETANUS TOXIN HEAVY CHAIN.
FT METAL 232 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 233 BY SIMILARITY.
FT METAL 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 236 246 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT DISULFID 438 466 INTERCHAIN.
FT DISULFID 1076 1092
SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF8ID CRC64;

Query Match 4.4%; Score 97.5; DB 1; Length 1314;
Best Local Similarity 20.0%; Pred. No. 20;
Matches 108; Conservative 78; Mismatches 154; Indels 201; Gaps 32;

QY 7 LVEYSLSTPPVARNRYTHVKEVREVNCSGIYEDP-PLEICKSLERRRDIIDLED-D 64
DB 827 LMOYIRANSKFI--GITELK---KLESKINKVSPPIPSYSKNLDC---WVDNEEDID 877
QY 65 VVAMTS-----DCDIYOTLGRYACKVSKSEKSPPIASVLMKDAIMV-----RLIH 113
DB 878 VILAKSTIILNDINDINDISIGFNS-----SVITYPDAQVGVINGKAIH 923
QY 114 AIYNQHNITCIHYDRKAPDTFKVAMNNLAKCFSNIF-----TASKEAV---EYAHIS 163
DB 924 LVNNESEVIVH---KAMD---LEYNDMFNNFTVSPFLRVPVSASHLEQYGTNEYSIIS 977
QY 164 RLQADLNCLS-----DLKSSIOKRY-----VINLGGDPEPLKSNFELVSEL----- 205
DB 978 SMKK--HSISIGSGMSVSLKGNLITLDSAGEVNOITFRDLPDKFNAYLANKWFTTI 1035
QY 206 --KRLGANK-----LETVRPNS--KLEFITYHHE----- 232
DB 1036 TNDRLSSANLYINGVLMSAEITGLGAIREDNNITLKLDRCNNNQYVSIDKFRIFCKAL 1095
QY 233 -----LR-----RVEYEVKLPFRINIKERAPHNI----- 258
DB 1096 NPKIEKLYTSYSTFLTRDPWGNPLRYDTETYLIPVASS--SKDVOAKNITDYMYLTNAP 1154
QY 259 -----QIFVGSAYEVLQSAFYKIFNNSIYODFFAMSKD-----TYSDEHEFWA--- 302
DB 1155 SYTNKLTNIYRRLNGKLFIIKRYTPNNEI--DSFVKSGDFIKLVSYNNNEHIYGPYK 1212
QY 303 -----TLIRV-----GIRGEISRSQDVSDLOSRTIKVKNMYE-----GEFYPS 343
DB 1213 DGAENLNDRLILVGNAPGIPLYKMEAVKLDL--KITYVOLKTIYDKRNASLGL----- 1266
QY 344 CTGSHRSYVCIFYGAELRWLIDGHWFAKFDKVPDPIILIKLAELKEQOARDITLPSSE 403
DB 1267 -VGTHNQOI---GNDRFDILIASNMVFNHLKDKI---LGC-----DMYEVPTD 1308
QY 404 K 404
DB 1309 E 1309

```

```

RESULT 15
YCF2_PINTH
ID YCF2_PINTH STANDARD; PRT: 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 244.6 KDA PROTEIN (ORF 2054).

```

```

GN YCF2.
GN Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Coniferales; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Nakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugiura M.;
RT "Loss of all nci genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii."
RT Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC - FUNCTION: NOT YET KNOWN.
CC - SIMILARITY: BELONGS THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, D17510; BAA04460.1;
DR InterPro: IPR001939;
DR Pfam: PF00004; AAA; 1;
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

```

```

Query Match 4.4%; Score 97.5; DB 1; Length 2054;
Best Local Similarity 21.9%; Pred. No. 36;
Matches 70; Conservative 55; Mismatches 86; Indels 109; Gaps 20;

QY 4 DIYVEYSLST--SPFYR-NRYTHVKDE-----VREYVN-CSGIYEDPELEIGK 48
DB 871 DLVR-SFLRLRLNLPVREKRYLSIEISITPLTKQIVNFEKNFCQPFKRSDEEN 929
QY 49 SLEIRRRDILEDDEDDVAMTSCDDIYOTLGRYAKVYS-----KEKSPPIASVLYHK 103
DB 930 FDQCFKR-----GFSSNGLIQT-RSYDDDLISEMFSKNEIEPRI-----Q 971
QY 104 DAIVERLTHAIYNOHNITCIHYDRKAPDTFKVAMNNLAKCFSNIFLASKLEAVEYAHIS 163
DB 972 DMEVTECKKNIVM-----DIDGRS-----TLSNSKREQNIYRISQIDSI---FS 1015
QY 164 RLQADLNCLSLSSSIOW-----KYVINLGGDPEPLKSNFELVSELKLNKAN----- 212
DB 1016 KW-----DLFTYMWFTSAMCKYIENML-----LDTLSEI-LHOSNFPVS 1057
QY 213 MLETKPPNSKLERETTYHELRVYEVYKLPFRINISKEAPPHNIQIFVGSAYEVLQSA 272
DB 1058 ILQNK-KHILKR-NILMEISHPLMERIDCKLPTNLIIN----- 1094
QY 273 FVKYIFNNSIYODFFAMSKD 292
DB 1095 --KFFFPNNKDFEPYCKD 1112

```

Search completed: August 2, 2001, 14:50:20  
Job time: 300 sec

Fri Aug 3 09:38:18 2001

us-09-645-192-2\_copy\_39\_453.rsp

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comphen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:45:50 ; Search time 25.93 Seconds  
(without alignments)  
1219.145 Million cell updates/sec

Title: US-09-645-192-2\_COPY\_39\_453  
Perfect score: 2194  
Sequence: 1 POKDIYLVESLSTSPFVRN.....DWITLPSEKLFMDNLTTS 415

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR-68:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	849.5	38.7	428	2	A46293 beta-1,3-galactosyl
2	711	32.4	400	2	A46297 beta-1,6-N-acetylgl
3	321.5	14.7	445	2	T21262 hypothetical prote
4	313.5	14.3	425	2	T20207 hypothetical prote
5	296	13.3	401	2	T24929 hypothetical prote
6	291	13.3	454	2	T21261 hypothetical prote
7	288.5	13.1	472	2	T22188 hypothetical prote
8	265	12.1	402	2	T24930 hypothetical prote
9	245	11.2	459	2	T24742 hypothetical prote
10	226	10.3	454	2	T25382 hypothetical prote
11	218.5	10.0	470	2	T32137 hypothetical prote
12	218	9.9	489	2	T24013 hypothetical prote
13	213.5	9.7	486	2	T21566 hypothetical prote
14	196.5	9.0	472	2	T33364 hypothetical prote
15	196	8.9	753	2	T24745 hypothetical prote
16	194.5	8.9	448	2	D85042 probable glycosyla
17	188	8.6	478	2	T27714 hypothetical prote
18	187	8.5	447	2	T00906 hypothetical prote
19	182.5	8.3	513	2	T20068 hypothetical prote
20	178	8.1	367	2	T28882 hypothetical prote
21	178	8.1	467	2	T21408 hypothetical prote
22	175.5	8.0	406	2	T36571 unknown protein F2
23	168	7.7	395	2	B86735 hypothetical prote
24	166.5	7.6	434	2	T51450 hypothetical prote
25	132	6.0	630	2	T02524 probable RING zinc
26	131	6.0	384	2	T02524 hypothetical prote
27	123.5	5.6	371	2	T21819 hypothetical prote
28	116	5.3	895	2	B64431 UDPglucose 6-den
29	108.5	4.9	661	2	S50734 WH3 protein - yea

30	108	4.9	702	2	E69498 hypothetical prote
31	106.5	4.9	1347	2	T41321 RB domain and Ank
32	104	4.7	2486	2	T48628 hypothetical prote
33	104	4.7	2485	1	H71621 serine/threonine-s
34	104	4.7	2628	2	S59413 probable membrane
35	103.5	4.7	588	2	H82877 hypothetical prote
36	102.5	4.7	415	2	T65223 heparin-binding fi
37	102.5	4.7	707	2	A54846 fibroblast growth
38	102.5	4.7	707	2	A38429 keratinocyte growt
39	102.5	4.7	769	2	S16236 fibroblast growth
40	102.5	4.7	820	2	S17295 fibroblast growth
41	102.5	4.7	821	1	TVMSBK fibroblast growth
42	102.5	4.7	821	1	TVHUF2 fibroblast growth
43	102.5	4.7	822	2	A45081 fibroblast growth
44	102.5	4.7	822	2	B54846 fibroblast growth
45	100.5	4.6	437	2	T23652 alpha-1,3-mannosyl

## ALIGNMENTS

RESULT 1  
A46293  
beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase  
C:Species: Homo sapiens (man)  
C:Date: 20-Oct-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: A46293  
R.Bierhuizen, M.F.: Fukuda, M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992  
A:Title: Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc)  
A:Reference number: A46293; M01D:93028457  
A:Accession: A46293  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-428 <BIB>  
A:Cross-references: GB:M97347; NID:9183440; PIDN:AAA35919.1; PID:9183441  
A:Experimental source: HL-60 cells  
A:Note: Sequence extracted from NCBI backbone (NCBIN:115900, NCBIN:115901)  
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	38.7%	Score	849.5;	DB 2;	Length	428;
Best Local Similarity	43.0%	Pred. No.	1.1e-56;			
Matches	161;	Conservative	70;	Mismatches	129;	Indels 41; Gaps 11;
QY	1	POKDIYLV-EYSLSTSPFVRNRYTHVKE-VR-----EYNGSGIYDEPLEI	46			
DB	14	PTKYEFVIVLSTLITFSVLR---IHOKPEFVSRYHLELAGENPSSDINCTKYLGQGVNEI	70			
QY	47	GK-----SLEIRRDIILEDODVVAAMTSDCIYOTLGGYAKTLVSKSEKSPPIASLV	100			
DB	71	QVKKLELLIVKFKRP--RWTPPDYINMTSDCSFIRKKRIYIEPASKDEAPPIASLV	128			
QY	101	VHKDAIWEPLIHAIYNQENIYCIHYDRKAPDTFKYAMNNLACSENFIFASKLEVEYA	160			
DB	129	VHNIEMIDRLALRAIYWPQNFYCVHDTKSEDEYLAVMGIACSFNVVASLESVYA	188			
QY	161	HISRLDNLCLSDILKSSIQMWYVNLICGDDPLKSNFELKLLKANGANLLETVKPP	220			
DB	189	SWSVQADLNCMDLYAMSNMKYILNLCMDPIKTNIETVARKLLINGENNLLETBRMP	248			
QY	221	NSKLEPFTYHEELRRVYEV--KLPIRTNISKADPPHNIQIVGSAFYVLSQAFKYIF	278			
DB	249	SHKEERL-----KKRYEVANGKL-TNIGTVMLPPLPLETPFGSAYFVVSSEYGVYI	300			
QY	279	NSNIVQPFAMSKDYSDEHFNATLIRVPGIVGEISRSAC-DVSDLOSRTFLVKNVYE	337			
DB	301	QNEKIQLMEADQTYSPDEYIMATIQRIPEVQSLASIKYDLSMDQAAFRVKNQYFE	360			
QY	338	GFF-----YPSCTGSHLSVCYIGAAELRWLKDGHAFANKFDSKYDPIIKCLAEKLE	392			
DB	361	GDVKGAPYPPCDGVHVSVCIFGAGDINMMLRKHLFAFKFVDVDFLFIQCLDEHLRH	420			



```

Db      67  ILAQITLKSQENAAQAEFGKIFGFOEPTSOEELERPLAYGMLVHGDFVQLSLLSATY 126
      117  NOAHNYCIHIDRKAPDTFKYAMNNLAKCFNSI--FIASKLEAVEYAHISLQADLNCISD 174
      127  QPOQFCFLAVDGNSSVFEIGLVRLSRGYNIGYFTIDELRMCGYELTSV--FQCVYD 183
      175  LKSSIQMKYVINLCGODFPLKSNFELVSELKLGANMLETVKPNKSLERFTYHHELR 234
      184  LAKIPSPMKKTFQYLSGVADALKSNLEMRILKLKNG-----SFAHEI- 225
      235  RVPREYVKLPRTNISKADPPHNIQIFVGSAYFLVSGAFYKIFPNNSIVODFFAMSKDYI 294
      226  -LPEEFYRLNKRKRPWSSPLPLYKTSL--SATF--SRKSANFVWNSERKYLEQIDFLRGTY 279
      295  SPDEHFNATLIRVPGI-----PGETISRAQDVSDL-QSKTRLVKWN-----NYE 337
      280  CADSLWATINGNKKVILKFSLEPMGCFDAKAWIHKKYRTRTGLKMGXGNOKINDNGYV 339
      338  GFERYP-----SCTGSHLRSCVCIYGAAELRWLIKDGHWTFANKFDSKVDPILIKLAEKL 390
      340  SRYQOYVRAFPVKCKGYYYRLSCVFGYVDLPNLINRHELVAAHKLYFSYQPAAFMCIVENS 399
      391  EEQ 393
      400  RQK 402

```

## RESULT 5

T24929

hypothetical protein T15D6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T24929

R:Dobson, R. submitted to the EMBL Data Library, November 1996

A:Accession: T24929

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-401 &lt;WIL&gt;

A:Cross-references: EMBL:Z83125; PIDN:CA805620.1; GSPDB:GN00019; CESP:T15D6.2

A:Experimental source: clone T15D6

C:Genetics:

A:Gene: CESP:T15D6.2

A:Map position: 1

A:introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3

```

Query Match      13.5%; Score 296; DB 2; Length 401;
Best Local Similarity 23.4%; Pred. No. 7.2e-15;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;

```

```

      23  THVKKEVYEVNCGSIYQEPLEIGKSLERRDIDLEDDDVYAMTSDCIIYOTLRGXA 82
      34  TDLDDDL--ELNGTNIL--QGLKNEKQLELINTKMI--EKLKNSIDRCHTVKSMRFEN 86
      83  QLVSKSEKSPPIAYSLVYHKDAIYVERLIHAIVNQHNYICIYHDKRAPDTFKYAMNNLA 142
      87  EYPLSEEEARFPLSYGLLYVKELSOVLPMLSSIYQPNQECYIAGVENSATFLILELTS 146
      143  KCFSNIFLASKLEAVEYAHISRLQADLNCISDLKSSIQMKYVINLCGODFPLKSNFELY 202
      147  DCFPN----KRPITWSEYETINSYVDCLFKLSHKNNMKYFQYLSGVADILPLKTNLEKV 201
      203  SELKRLNGANMLETVKPNKSLERFTYHHELRVPEYVKLPRTNISKADPPHNIQIFV 262
      202  RLKLSKNGANVEIKYENRR--LQNGHEESPLPLFK 237
      263  GSAYFLVSGAFYKIFPNNSIVODFFAMSKDYISPDHFNATLIRVPG--IPGELSRSAQ 319
      238  SSLSLSIPRKANYLASSSIPOOLLLEFLNRTVAVADEGFWGLGNKGLDVPVGSIN---- 293

```

```

      320  DVSDLSQKTRLVKWNYYEGFYPSCGTGSHLRSCVCIYGAAELRWLIKDGHWTFANKFDSKV 379
      294  -----FEEHQWFESGCHNNMKDSCVFGIGVSNLLQAKALVAHKLYLTSE 340
      380  PILIKCL 386
      341  PEAYFCL 347

```

## RESULT 6

T21261

hypothetical protein F22D6.12 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T21261

R:Wilkinson, J. submitted to the EMBL Data Library, April 1996

A:Reference number: 219397

A:Accession: T21261

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-454 &lt;WIL&gt;

A:Cross-references: EMBL:Z71262; PIDN:CA935816.1; GSPDB:GN00019; CESP:F22D6.12

A:Experimental source: clone F22D6

C:Genetics:

A:Gene: CESP:F22D6.12

A:Map position: 1

A:introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3

```

Query Match      13.3%; Score 291; DB 2; Length 454;
Best Local Similarity 24.1%; Pred. No. 2e-14;
Matches 99; Conservative 73; Mismatches 174; Indels 64; Gaps 14;

```

```

      17  FVRNRYTHVKDEV---RYEVNCGIYQEPLEIGKSLERRDIDLEDDDVYAMTSDC 73
      40  FLKSICTTASDSYLLDNMEINCSNII--KGYKTNEKIDIMHDI--EQELFSCINKCQ 94
      74  IYOTLRGYAQKLVSKSEKSPPIAYSLVYHKDA-----INVERLIHAI 115
      95  TLKTLFRNTNPMASAEKHPPLSYGLMYKLDLPQYPARMFILKHELEINLQVFLLSI 154
      116  YNQHNYICIYHDKRAPDTFKYAMNNLAKCFNSIFISKLEAVEYAHISRLQADLNCISL 175
      155  YHPQNEYCIAVGNSAPIFQNLREVSTCFSNVHF--MKRPPISWGSHEIIDSYYDLEFL 213
      176  LKSSIQMKYVINLCGODFPLKSNFELVSELKLGANMLETVKPNKSLERFTYHHELR 235
      214  SLETDRIYQYLSGVDPILKTLDEYQILKLNQTSNVEIT--NYQARLIGKNE--- 267
      236  VPEYVKLPRTNISKADPPHNIQIFVGSAYFLVSGAFYKIFPNNSIVODFFAMSKDYIS 295
      268  -----NESP---LPLFKSSLSAIIIPRKANQULASSNTARKLLEFLNTEI 309
      296  PDEHFNATLIRVP--GIPGEI-SRSAQDVSDQSK--TRLVKWNYY---EGFFYPSCTG 346
      310  ADEGFWGLTGNKDQFPISSINSKDMMEYRDQNNIINFNTDQMSYIISRDQIMPELCK 369
      347  SHLR-SVCIVYGAELRWLIKDGHWTFANKFDSKVDPILIKLAEKLEEQR 395
      370  NYMKDSCVFGIGVPLRLTFSKALVAHKFYLKSEPEAYFCL--LKEHNR 416

```

## RESULT 7

T22188

hypothetical protein f44F4.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T22188

R:Colles, L. submitted to the EMBL Data Library, September 1994

A:Reference number: Z19528

A:Accession: T22188



A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-472 <WIL>  
 A:Cross-references: EMBL:Z37092; PIDN:CA85457.1; GSPDB:GN00020; CESP:F44F4.6  
 A:Experimental source: clone F44F4  
 C:Genetics:  
 A:Gene: CESP:F44F4.6  
 A:Map position: 2  
 A:introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

Query Match 13.1%; Score 288.5; DB 2; Length 472;  
 Best Local Similarity 22.0%; Pred. No. 3.3e-14;  
 Matches 91; Conservative 78; Mismatches 175; Indels 69; Gaps 13;

QY 40 EDEPLEIGKSL-IRRDI-----IDLEDDVAMTSDCDIYQTLRGYAKLVSK 88  
 Db 63 ENPLECNSVINGTENRRKISRARQMDWHEHEIFNSRWVCSIDKXFNTRIPSSR 122  
 QY 89 EKSFPILASLVYHKDAIMVERLIHAIYNQHNICYHYDRKAPDFKAMNLLACFSN- 147  
 Db 123 EEAEYPLAAGLVVKTIVQVLTQMSLFYOPQHFCTTDDQSEKYSIQALPSCFFNM 182  
 QY 148 -IFIAKLEAVEYAHISRLOADNCLSDLLKSSIQKRYVINLGGDFPLKSNFELVSELK 206  
 Db 183 HAFIG--EPSQMGSGILKNVYTCFNMLSKSKQKKYQYLSGTDLPRTNLEWYRIK 239  
 QY 207 KLNQANMLEVYKPPNSKLEFTHHRLRVYEVYVVLPIRTINISKAPPHNIOIFGSAV 266  
 Db 240 ALNGS-----MNTDVSTF---FVDRP-----KNNEGVLP--NPVYKSSS 275  
 QY 267 FVLSQAFVKYIFNNSIVODFFANSKDTPSPDEHFMATLIRVP--IPGEISRSQDVSD 323  
 Db 276 VVVPREBADDLISPRQKTLKLTWIPDESFMTJVGSPALLPVPST--RVNDILM 333  
 QY 324 LQSKTRL-----VKNNYIEG---FYSCGSHLRSCVITGAAELRMLKDGHM 369  
 Db 334 LRKFKFLRPYENTVNSIGSTYGRQVWGMQKCEGKVKDFSCVGVGVDIEBIRPEL 393  
 QY 370 FANFDSKVPDILIKLAELKLEQQ-----RDMITLPSEKLFPMRNL 412  
 Db 394 VAKHLIEFQPAFMCLKLEVRKRSLSPAHLFSANYSQMPYVELXQKAIT 446

RESULT 8  
 T24930  
 hypothetical protein T15D6.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T24930  
 R:Dobson, R.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19956  
 A:Accession: T24930  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-402 <WIL>  
 A:Cross-references: EMBL:Z83125; PIDN:CA805621.1; GSPDB:GN00019; CESP:T15D6.3  
 A:Experimental source: clone T15D6  
 C:Genetics:  
 A:Gene: CESP:T15D6.3  
 A:Map position: 1  
 A:introns: 55/3; 85/2; 121/3; 161/3; 237/3; 262/3

Query Match 12.1%; Score 265; DB 2; Length 402;  
 Best Local Similarity 24.7%; Pred. No. 1.6e-12;  
 Matches 93; Conservative 64; Mismatches 157; Indels 62; Gaps 13;

QY 23 THVDEVEYVNGSIVOEPLIGKSLERRDIIDLEDDVYAMTSDCDIYQTLRGYA 82  
 Db 44 TDLDDL--QINCINILGFKNE--NTELELINTKMT--EKKMMNSTDRQTLTSMFRN 96

QY 83 QKLVSKKEKSEFPPLASLVYHKDAIMVERLIHAIYNQHNICYHYDRKAPDFKAMNLLA 142  
 Db 97 KVPISSEEARFPLSFGLLVYKELSOVLFLLSSIVOPQNEFCIVGENSAPAFILLKEEA 156  
 QY 143 KCFENIFIAKLEAVEYAHISRLOADNCLSDLLKSSIQKRYVINLGGDFPLKSNFELV 202  
 Db 157 NCFPN-----KRPPIKNGSEIINSYGCLEFSLHLSKMDKRYOYLSGVDIPLKTNEV 211  
 QY 203 SELKLNQANMLEVYKPPNSKLEFTHHRLRVYEVYVVLPIRTINISKAPPHNIOIF 262  
 Db 212 RILKRLNG-----TVNIGIS-----TYEDRL-----LNGKXRTESP--LPLFK 247  
 QY 263 GSATFVLSQAFVKYIFNNSIVODFFANSKDTPSPDEHFMAT-----LIRVPGEISR 316  
 Db 248 SLSLIRKAAANYLSSSVPOQLLEFLRTTVADEGFMTGLGNKDLFVNPSPFNNDP 307  
 QY 317 SAQVSDLOSKTRLVKW-----NYEGFFYPSCTGSHLRSCVITGAAELRMLKDGHM 370  
 Db 308 LTNGWGVVSRHQ--WVESECHNYMRD-----RS-CVFGIGDVNLMKSRALV 353  
 QY 371 ANKFDKVPDILIKCL 386  
 Db 354 AHKLYTESDEPAFCL 369

RESULT 9  
 T24742  
 hypothetical protein T09E11.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
 C:Accession: T24742  
 R:McLay, K.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19930  
 A:Accession: T24742  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-459 <WIL>  
 A:Cross-references: EMBL:Z81147; PIDN:CA803536.1; GSPDB:GN00019; CESP:T09E11.9  
 A:Experimental source: clone T09E11  
 C:Genetics:  
 A:Gene: CESP:T09E11.9  
 A:Map position: 1  
 A:introns: 39/1; 78/3; 204/1; 291/1; 338/3; 371/3; 417/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 11.2%; Score 245; DB 2; Length 459;  
 Best Local Similarity 22.8%; Pred. No. 6.2e-11;  
 Matches 106; Conservative 74; Mismatches 202; Indels 82; Gaps 17;

QY 5 IYLVESLSTSPFVNRNRTHVNDVRYVNCGLY--EPEPLEIGKSLERRDIIDLE 62  
 Db 25 VYLLMFS-SKSLFGRN-----EONSVALTNLNENYORAIINELRPDFAKTES 72  
 QY 63 -----DDVYAMTSDCDIYQTLRGYAKLVSKSEKSF----- 94  
 Db 73 KIEHFYSKYRRRSETHAVDCGRILSGCKDYQTVSG--ENRIPAVENPNPDMSCSAVMD 130  
 QY 95 -----IAYLVYHKDAIMVERLIHAIYNQHNICYHYDRKAPDFKAMN 139  
 Db 131 RIIPHDHILRPKNGVAFKRIYKDEYERQVQMSYHPONSFCALDKKAPTRFKKMR 190  
 QY 140 NLAQFSNFIASKLEAVEYAHISRLOADNCLSDLLKSSIQKRYVINLGGDFPLKSNF 199  
 Db 191 AMAACIPVLLLPDPEIDISNGHNVLNCLRALINKP--GMNYAMLLQNHDLITKSYV 249  
 QY 200 ELVSEKLNQANMLEVYKPPNSKLEFTHHRLRVYEVYVVLPIRTINISKAPPHNIO 259  
 Db 250 ELEQVYEWLGAGANDVELL--PEAQRLDEENFKWD---PSLAMPFDSKVPDETILNEKIK 304  
 QY 260 IFVGSAYFVLSQAFVKYIFNNSIVODFF--AMSKDYVSPDEHFMATL--IRVPGIGELISR 316

Db 305 FSKGVQSGSKRAAVDMTRKMYLSTYIDQWNGRWGVDEMLISSQISAFGLMPGHF-- 362  
 QY 317 SADDVSDLSKRTLYKMYEYGFYSCGSHLR-SVCITYGAELFWLJKDGHWFANKPD 375  
 Db 363 TDGCLGKGPFFMOWNFDESY---CASKYRHNVCILGIEFLYSVASFPILMNKML 419  
 QY 376 SKVDPILIKCLAELEQ---ORDWITLPESEKLFMDRLTTTS 415  
 Db 420 PPDNSIIECTAELLYNRTFMQNDH---PLEBEY-KMWTVS 459

RESULT 10  
 T2382  
 hypothetical protein T27F6.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
 C:Accession: T2382  
 R:Dobson, R.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z20026  
 A:Accession: T2382  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-454 <MIL>  
 A:Cross-references: EMBL:282060; PIDN:CAB04881.1; GSPDB:GN00019; CESP:T27F6.1  
 A:Experimental source: clone T27F6  
 C:Genetics:  
 A:Gene: CESP:T27F6.1  
 A:Map position: 1  
 A:Introns: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 10.3%; Score 226; DB 2; Length 454;  
 Best Local Similarity 24.4%; Pred. No. 1.7e-09;  
 Matches 80; Conservative 56; Mismatches 158; Indels 34; Gaps 12;

QY 84 KLVKEEKESEF---IAYSLVHKDAIMVERLHAIVQHNHCYIHRYKRAPDFTKYAMN 139  
 Db 110 RILSRDVLRLNGVAFARVYVMDYELIEKHVMSYHPQNSCFALDKAAKEFERMQ 169  
 QY 140 NLAKFSNIF-----ASKLEAVEYAHISRLQDLNCLSDLKSSIQWKYVNLGC 190  
 Db 170 AMASCLPVLILLGRFEPKPIHDSVDSHGHTNL-AHYNCLRALINKP-GWNPAILLON 227  
 QY 191 ODFPKSFEELVSELKKNLGMLETYKPPSKLERETYNHLEKRYEYKLPRTNIS 250  
 Db 228 HDLTKSYELEKIFNMVGANDY-AIRPELGRDK-----KHFKMDMSLKLFENES 279  
 QY 251 KEAP---PHNIQIVGSAFYVLSQAFYKIFNNSIVODFF-AWSKDTYSPDEHFMAT--L 304  
 Db 280 EIDVYIINTLTILKFAKGAOSSLISRAAVDMTRVDTLTFTIDQWNGRWGVDEQFTQAFQI 339  
 QY 305 IRVGIPEIS-RSAQDVSDLSKRTLYKMYEYGFYSCGSHLR-SVCITYGAELRW 362  
 Db 340 SDFLGMPEHFDKCIKGIITEGITRFQWTHGD---QSKCAKSKSHGICIMGIEHLSM 396  
 QY 363 LKDGHWFAKFDKVDPIILKCLAEKL 390  
 Db 397 MAKSEHLMFNKVLPLFDYSIIECTAELL 424

RESULT 11  
 T2137  
 hypothetical protein R07C3.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 20-Jun-2000  
 C:Accession: T2137  
 R:Lamar, B.; Kramer, J.  
 submitted to the EMBL Data Library, July 1997  
 A:Description: The sequence of C. elegans cosmids R07C3.  
 A:Reference number: Z21125  
 A:Accession: T2137

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-470 <LAM>  
 A:Cross-references: EMBL:AF016686; PIDN:AAB6233.1; GSPDB:GN00020; CESP:R07C3.3  
 A:Experimental source: strain Bristol N2; clone R07C3  
 C:Genetics:  
 A:Gene: CESP:R07C3.3  
 A:Map position: 2  
 A:Introns: 18/1; 78/3; 189/1; 276/1; 319/3; 359/2; 401/3; 441/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 10.0%; Score 218.5; DB 2; Length 470;  
 Best Local Similarity 21.2%; Pred. No. 6.5e-09;  
 Matches 90; Conservative 85; Mismatches 188; Indels 61; Gaps 17;

QY 22 YTHVKDEVR-YEVNC-SGIYEQEP-----LEIKSLERIRRDII--DLEDDVYA 67  
 Db 39 YKRIEALRTVDRCSEEGGFYCKRPETQHVDCGRVLDGDKIQLRFAFYQLSLGSGSNRP 98  
 QY 68 MTSDDDIQTLRGVAKIYKSEKSEKSPY----AYSLVHKDAIMVERLHAIVQHNHY 122  
 Db 99 LENSEFLNLTGSAIKSRITIPKTAQFKPLKNGTARARIYPADYETIEKQVQASHYPQNVF 158  
 QY 123 CIHYDKAPDTEFKVAMNLAQCFNIFLASKLEAVEYAHISRLQDLNCLSDLKSSIQW 182  
 Db 159 CRAIDANSSAEQRMKKKLEQCLPNNVYLPVTEYSYDSGHNINLAHYCMKK-LESIRGW 217  
 QY 183 KYVINLCCGDFPLKSNFELVSELKKNLGMLETYKPPSKLERETYNHLEKRYEYK 242  
 Db 218 GYIMLQNHNDVITKSYELDRIFELLGVN-----DVESKRIPERRKKHLMWDLKS 269  
 QY 243 LPIRTNISKKEAPPHNIQIVGSAFYVLSQAFYKIFNNSIVODFF-AWSKDTYSPDEHFM 301  
 Db 270 LKLFNRDSKNS-DYELFISGSHSLSAAYKWLVEVDLSTFIDQWNGRWGVDEQFTI 328  
 QY 302 AFLIRVPI--PGEISRSA-QDVSDLSKRTLYKMYEYGFYSCGSHLRASCIYGA 357  
 Db 329 STFQMSPDLMKMGHFNEDCIHNDTAVITISRLARWPYILDRAAHCACTGVRHDCITGI 388  
 QY 358 AELRWLIKDGHWFAKFDKVDPIIL-----IKCLAELEQQRDMWITLPESEKLFM 407  
 Db 389 EDPRAI-----SKF-----PILNMLPAFDYSIIECTAELL--HNRYYLGQVQDKI-- 431  
 QY 408 DNNL 411  
 Db 432 EKNL 435

RESULT 12  
 T24013  
 hypothetical protein R07B7.6 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T24013  
 R:Harris, B.  
 submitted to the EMBL Data Library, July 1996  
 A:Reference number: Z19830  
 A:Accession: T24013  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-489 <MIL>  
 A:Cross-references: EMBL:275955; PIDN:CAB00115.1; GSPDB:GN00023; CESP:R07B7.6  
 A:Experimental source: clone R07B7  
 C:Genetics:  
 A:Gene: CESP:R07B7.6  
 A:Map position: 5  
 A:Introns: 42/3; 80/2; 235/3; 287/1; 372/2; 416/1

Query Match 9.9%; Score 218; DB 2; Length 489;  
 Best Local Similarity 22.0%; Pred. No. 7.5e-09;  
 Matches 97; Conservative 74; Mismatches 197; Indels 72; Gaps 18;



hypothetical protein T09E11.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T24745

R:McClay, K.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19930

A:Accession: T24745

A:status: preliminary; translated from GB/EMBL/DBJ

A:molecule type: DNA

A:Residues: 1-753 <WIL>

A:Cross-references: EMBL:Z81147; PIDN:CAM03539.1; GSPDB:GN00019; CESP:T09E11.6

A:Experimental source: clone T09E11

C:Genetics:

A:gene: CESP:T09E11.6

A:map position: 1

A:introns: 34/2; 54/2; 191/3; 275/3; 315/3; 359/3; 380/3; 403/3; 426/3; 530/1; 617/1; 66

#### Query Match

8.9%; Score 196; DB 2; Length 753;

Best Local Similarity 21.5%; Pred. No. 6.3e-07;

Matches 88; Conservative 72; Mismatches 195; Indels 54; Gaps 14;

```
QY 7 LVEYSLS-----TSPVRNRYTHVKEVR--YEVNCSGIYQEP---LEIGKSLER 53
DB 366 LSESTLNCCEGPTKODLHHHTDIDAVYREGRFADPKRGSYRRRPETAHVDCGRILAGD 425
QY 54 RROIDLEDDDVAMTSDCDIYOTLRGAKIVSKEEK-----SFLAYSLVYHKDAIMV 108
DB 426 KPYLOSITGTRKRVIVENCNINNSCKAIRSILPSNDNIIRPLKHGIAFARIYKDYEFI 485
QY 109 ERLIHAIYNQHNIIYCIHYDRKAPDPTFKVAMNNLAKCFSNIFIAASKLEAVEYAHISRLQAD 168
DB 486 EKQYVSFHPQNAFCFVIDINASEEFKRRALACMPNVIYLADEDPVYSSGHNVLVH 545
QY 169 LNCISDLKSSIQKRYINLCGQDFPLKSNPELVSELKINGANMLETVPKPNKLERFT 228
DB 546 NKCKLALDIP-GNNYALLQNHDLIMKSYEMEQITEWLOGANDIFVT-----593
QY 229 YHHELRRVPEYVKL-PI-----RTNISKAPPHNIQIFVGSAYFVLQAIFYKYLENN 280
DB 594 --HEIGRYDVAKKLKWDPMSTKLFTINETEMDKLLLTTPMKIYKGVWHSLSRASVEAMFQK 651
QY 281 SIQDDEF-AMSKDIYSPDEHFWATLIRP--GIRGEIS-RSADQVSDLOSRTLYKWNYY 336
DB 652 LDPSTLFMHQLNQGRYGVDEQYFPIIQANAEFGMPGHHFTDECLQOGKTTERTTITIALM---708
QY 337 EGFFYP--SC-TGSHLRSCVCIYGAELRWLIKDGHWFAFKPDSKVPDIL 382
DB 709 ---VPESKCDTNTMRHVAVCITGLEHFGAVASFTHLMFNKYSLSLDELI 753
```

Search completed: August 2, 2001, 14:45:51

Job time: 137 sec



XX MO200114535-A2.  
 PN  
 XX  
 PD 01-MAR-2001.  
 XX  
 XX 24-AUG-2000; 2000MO-DK00469.  
 PF  
 XX 24-AUG-1999; 99US-0150488.  
 PR  
 XX (SCHW/) SCHWIENTEK T.  
 PA (CLAU/) CLAUSEN H.  
 XX  
 PI Schwiientek T, Clausen H;  
 XX  
 DR WPI; 2001-226615/23.  
 DR N-PSDB; AAS00045.  
 XX  
 PT New C2GNT3 polypeptides and nucleic acids encoding the polypeptides  
 PT useful for treating conditions mediated by a C2Gnt3 polypeptide, e.g.,  
 PT thymus-related disorders, cancers, tumours, immunosuppression  
 PS  
 PS Claim 17; Fig 1; 97p; English.

The sequence represents Human UDP-N-acetyl-glucosamine:Galactose-beta1,  
 3-N-acetylglucosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase  
 (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or  
 C2Gnt3). C2Gnt3 and nucleic acids encoding it are useful in the  
 preparation of compositions for treating a conditions mediated by C2Gnt3,  
 particularly a thymus-related disorder. C2Gnt3, nucleic acids  
 encoding it and antibodies against it may also be used for in vitro  
 purposes related to scientific research, DNA synthesis and manufacture of  
 vectors, in the prognostic and diagnostic evaluation of conditions  
 associated with altered expression or activity of C2Gnt3 or conditions  
 requiring modulation of C2Gnt3, as well as in monitoring conditions by  
 detecting and localising the DNA and protein. Disorders such as tumours  
 (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the  
 breast or cervix), hypochlorid, hyperactivity, atrophy, enlargement of  
 thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression  
 acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,  
 sepsis, wound healing, acute and chronic infection, cell-mediated or  
 humoral immunity, or PH1/TH2 imbalance, may be treated using these protein  
 or nucleic acid. The antibodies may be used to screen potential  
 therapeutic compounds to determine their effects on a conditions such as  
 thymus-related disorder or cancer, to determine the level of C2Gnt3  
 expression in cells genetically engineered to produce C2Gnt3, or to  
 detect and quantify polypeptides in a sample to determine their role in a  
 particular cellular events or pathological states and to diagnose and  
 treat such pathological states.

XX  
 XX Sequence 453 AA;

Query Match 100.0%; Score 2194; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 2,9e-190;  
 Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 POKKITYLVESLSIPVVRNRTYVKKDEVYEVNCSGIYQEPLEIKSLFIRRDIDL 60  
 Db 39 PPKKITYLVESLSIPVVRNRTYVKKDEVYEVNCSGIYQEPLEIKSLFIRRDIDL 98  
 XX  
 XX 61 EDDVYVMTSDCDYQTLRGYAKLVGKEKSEPIAYSLVNHKDALVERLHAIIYQNH 120  
 Db 99 EDDVYVMTSDCDYQTLRGYAKLVGKEKSEPIAYSLVNHKDALVERLHAIIYQNH 158  
 XX  
 XX 121 IYCHYDRKAPDTKVMNNIAKFSNITFASKIENAEVYAHISLQADNLNCLDLKST 180  
 Db 159 IYCHYDRKAPDTKVMNNIAKFSNITFASKIENAEVYAHISLQADNLNCLDLKST 218  
 XX  
 XX 181 QMKYVINLCGDPFLKSNFELVSELKLNANMLETVKPPNSKLEERTYVHHELRVPEY 240  
 Db 219 QMKYVINLCGDPFLKSNFELVSELKLNANMLETVKPPNSKLEERTYVHHELRVPEY 278  
 XX  
 XX 241 VKLPFRINISKEAPPHNIQIFVGSATYVLSQAFVKYIFNNSTYVDFAWSKDTYSPDEH 300

Db 279 VKLPFRINISKEAPPHNIQIFVGSATYVLSQAFVKYIFNNSTYVDFAWSKDTYSPDEH 338  
 XX  
 XX 301 WATLIRVPGIGETISRSADVSDLOSRTPLVKNMYGCFYPSCTGSHLSVCYIAAEL 360  
 Db 339 WATLIRVPGIGETISRSADVSDLOSRTPLVKNMYGCFYPSCTGSHLSVCYIAAEL 398  
 XX  
 XX 361 RWLIRKGHWFAKRPDSKYDPIILKICLAEKEBOORDMTLPSEKLFMDRLTTTS 415  
 Db 399 RWLIRKGHWFAKRPDSKYDPIILKICLAEKEBOORDMTLPSEKLFMDRLTTTS 453

# RESULT 2

ID AAB18999 standard; Peptide: 406 AA.

AC AAB18999;

DT 08-FEB-2001 (first entry)

DE A core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region.

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;  
 KW Cancer; cardiovascular disorder; inflammatory disorder; asthma;  
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;  
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;  
 KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;  
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;  
 KW diverticulitis; ulcerative colitis.

OS Homo sapiens.

PN CA2296936-A1.

PD 03-AUG-2000.

PF 03-FEB-2000; 2000CA-2296936.

PR 03-FEB-1999; 99US-0118674.

XX (GLYC-) GLYCDESIGN INC.

XX Korszak B, Lew A;

XX WPI; 2000-594746/57.

PT New nucleic acid molecules of core 2  
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new  
 PT compositions for treatment of disorders mediated by the enzyme  
 PT including cancer, cardiovascular and inflammatory disorders.

XX Claim 3; Page 53; 66pp; English.

The present sequence is derived from a human core 2  
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.  
 CC The polypeptide can be used to treat diseases and disorders, such as  
 CC cancer, cardiovascular disorders and inflammatory disorders including  
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,  
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome  
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as  
 CC atherosclerosis and clotting can also be treated. The polypeptides of  
 CC the invention are predominantly expressed in gastrointestinal tissue  
 CC (stomach, colon, intestine, testis) and are elevated in cancer.  
 CC Gastrointestinal disorders that may be prevented or treated include  
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis  
 CC and ulcerative colitis. The antibodies may be used in  
 CC immuno-histochemical analysis to detect the novel polypeptide and to  
 CC localize it to particular cells and tissues and to specific subcellular  
 CC locations and to quantitate the level of expression.

XX Sequence 406 AA;

Query Match 39.3%; Score 862; DB 21; Length 406;  
 Best Local Similarity 44.5%; Pred. No. 9.2e-70;  
 Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps 8;

QY 13 STSPFYRN-RYTHVKDEVRVENCSGIY--EQEPL--EIGKSLERIRDDIILEDDDVYA 67  
 DQ 15 SSGYQYRNLLYFKLPKATKINSYGVTRGQDEAVQALINLEVKKR-epftdthyls 73  
 QY 68 MTSDDCIYQTLRGYAOGLVSKSEKSPPIAYSLVYHKDAIWERLHAITYNQHNICYIHD 127  
 DQ 74 ITRDCENKAEKRFIQLPISKSEVEFIAYSMVHEKLENFERLIRAVAPQNIYCVHVD 133  
 QY 128 RKAPDTEKVMNNLAKCFNSIFASKLEAVEYAHISRLQADINCLSDLSKSIOMKYVIN 187  
 DQ 134 EKSPCTKEAVKALISCFNVIASKLIVRYVYASWSTVQDINMEDLIQSSVPMKYFIN 193  
 QY 188 LCGQDFPLKSNFELVSELKRLGANKMLETVPKNSKLERFTYHHELRVPEYVKLPIRT 247  
 DQ 194 TCGTDFPKSNAEMVQALKNLNGRNSMESEVPKHKETRWKYHFEVYR--dtlhl--t 247  
 QY 248 NISKEAPPNHIQIFVGSAYFVLSQAFYKIFNNSIVODPFAMSKDTYSPDEHFMATIRV 307  
 DQ 248 NKKKDPPIYNTLTGTNAIVASRDIYVHLKPKSGQLIEWVXDLYSPDEHFMATIRV 307  
 QY 308 PGIPGEI-SRSADVPDLSQKTRLVKWNYYEGF----YPSCTGSHLRSCYIYGAELR 361  
 DQ 308 RWPMSGVNHPKYDISMTSLARLVKVGHEGDIKQAPYPCSGIHQRAICVYGAGDIN 367  
 QY 362 WLKDGHWFMANKPDSKYDPLILICIAEKL 390  
 DQ 368 WMLQNHHLANKFKDPKVDNAIGCLEEYL 396

## RESULT 3

AA18995 ID AA18995 standard; Protein: 438 AA.

AA18995; AC

08-FEB-2001 (first entry) DT

A core 2 beta-1,6-N-acetylglucosaminyltransferase. DE

Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;  
 cancer; cardiovascular disorder; inflammatory disorder; asthma;  
 rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;  
 septic shock; adult respiratory distress syndrome; ARDS; cancer;  
 platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;  
 clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;  
 diverticulitis; ulcerative colitis. KM

Homo sapiens. OS

CA2296936-A1. PN

03-AUG-2000. PD

03-FEB-2000; 2000CA-2296936. PF

03-FEB-1999; 990S-0118674. PR

(GLYC-) GLYCDESIGN INC. PA

Korczak B, Lew A. PI

WPI: 2000-594746/57. DR

N-PSDB: AAA96569. XX

New nucleic acid molecules of core 2  
 beta-1,6-N-acetylglucosaminyltransferase useful for providing new  
 compositions for treatment of disorders mediated by the enzyme  
 including cancer, cardiovascular and inflammatory disorders. PT  
 XX

PS Claim 3; Page 50-51; 66pp; English.

XX The present sequence represents a human core 2  
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.  
 CC The polypeptide can be used to treat diseases and disorders, such as  
 CC cancer, cardiovascular disorders and inflammatory disorders including  
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,  
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome  
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as  
 CC atherosclerosis and clotting can also be treated. The polypeptides of  
 CC the invention are predominantly expressed in gastrointestinal tissue  
 CC (stomach, colon, intestine, testis) and are elevated in cancer.  
 CC Gastrointestinal disorders that may be prevented or treated include  
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis  
 CC and ulcerative colitis. The antibodies may be used in  
 CC immuno-histochemical analysis, to detect the novel polypeptide and to  
 CC localize it to particular cells and tissues and to specific subcellular  
 CC locations and to quantitate the level of expression. XX

Sequence 438 AA.

Query Match 39.3%; Score 862; DB 21; Length 438;  
 Best Local Similarity 44.5%; Pred. No. 1e-65;  
 Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps 8;

QY 13 STSPFYRN-RYTHVKDEVRVENCSGIY--EQEPL--EIGKSLERIRDDIILEDDDVYA 67  
 DQ 47 SSGYQYRNLLYFKLPKATKINSYGVTRGQDEAVQALINLEVKKR-epftdthyls 105  
 QY 68 MTSDDCIYQTLRGYAOGLVSKSEKSPPIAYSLVYHKDAIWERLHAITYNQHNICYIHD 127  
 DQ 106 ITRDCENKAEKRFIQLPISKSEVEFIAYSMVHEKLENFERLIRAVAPQNIYCVHVD 165  
 QY 128 RKAPDTEKVMNNLAKCFNSIFASKLEAVEYAHISRLQADINCLSDLSKSIOMKYVIN 187  
 DQ 166 EKSPCTKEAVKALISCFNVIASKLIVRYVYASWSTVQDINMEDLIQSSVPMKYFIN 225  
 QY 188 LCGQDFPLKSNFELVSELKRLGANKMLETVPKNSKLERFTYHHELRVPEYVKLPIRT 247  
 DQ 226 TCGTDFPKSNAEMVQALKNLNGRNSMESEVPKHKETRWKYHFEVYR--dtlhl--t 279  
 QY 248 NISKEAPPNHIQIFVGSAYFVLSQAFYKIFNNSIVODPFAMSKDTYSPDEHFMATIRV 307  
 DQ 280 NKKKDPPIYNTLTGTNAIVASRDIYVHLKPKSGQLIEWVXDLYSPDEHFMATIRV 339  
 QY 308 PGIPGEI-SRSADVPDLSQKTRLVKWNYYEGF----YPSCTGSHLRSCYIYGAELR 361  
 DQ 340 RWPMSGVNHPKYDISMTSLARLVKVGHEGDIKQAPYPCSGIHQRAICVYGAGDIN 399  
 QY 362 WLKDGHWFMANKPDSKYDPLILICIAEKL 390  
 DQ 400 WMLQNHHLANKFKDPKVDNAIGCLEEYL 428

## RESULT 4

AA194492 ID AA194492 standard; Protein: 438 AA.

AA194492; AC

19-SEP-2000 (first entry) DT

Human C2/4Gnt protein. DE

Human C2/4Gnt protein. OS

Human C2/4Gnt: UDP-N-acetylglucosamine: O-glycan biosynthesis; KM

O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer. XX

Homo sapiens. XX  
 Key Location/Qualifiers  
 FT Domain 10..27 /note= "putative transmembrane domain" FT





QY 188 LCGDPLKSNFELVSEIKLNGANMLETVPKPSKLERETVHHLELRVREYVKLPRT 247  
 Db 226 tctgdfipkksnaevqaklmngnsmesevppkhetrvkyhevr---dclhl---t 279  
 QY 248 NISKEAPPHNIQIFVGSAYFVLSQAFVYIFNNISIVODFFAMSKDYSPPDEHFWATLIRV 307  
 Db 280 nkkddppnylmtftgnayivasdfvghvklpkbsqgllewvxdtyspdehlwatlqra 339  
 QY 308 PGIREFI-SRGAQVSDLOSKTRLYKKNYEGFF-----YPSCTGSHLRSCVCTIGAAELR 361  
 Db 340 rmpgsvpnbpkkydidsmtslarlvkwsqhegdldkgapapcsqllnqralcvygsqdl 399  
 QY 362 WLKDGHWFAKFKDSKVDPIIKLAERL 390  
 Db 400 wmlqnhllankfdpkyddnalqcleeyl 428

RESULT 6  
 AAB54344  
 ID AAB54344 standard; Protein: 465 AA.  
 XX AAB54344;  
 AC AAB54344;  
 DF 09-MAR-2001 (first entry)  
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.  
 XX Human: pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KM detection; diagnosis; identification; cytosolic; neuroprotective;  
 KM neoplastic; immunomodulatory; relaxant; contractile; gynaecological;  
 KM antinflammatory; cardiant; gene therapy; chromosome mapping;  
 KM linkage analysis; tissue identification; tissue typing; forensic;  
 KM neural; immune system; muscular; reproductive; gastrointestinal;  
 KM pulmonary; cardiovascular; renal; proliferative.  
 XX Homo sapiens.  
 OS WO200055320-A1.  
 XX 21-SEP-2000.  
 PD 08-MAR-2000; 2000WO-US05989.  
 PF 12-MAR-1999; 99US-0124270.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM;  
 XX MPI; 2000-579444/54.  
 DR N-PSDB; AAC99109.  
 XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX Claim 11; Page 1237-1239; 1379pp; English.

CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.  
 XX

SO Sequence 465 AA:  
 Query Match 39.3%; Score 862; DB 21; Length 465;  
 Best Local Similarity 44.5%; Pred. No. 1 le-69;  
 Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps 8;

QY 13 STSPFVRN-RYTHVKDEYREYVNCSGIY--EQPPL--ETGKSLERRDITLEDDEYVA 67  
 Db 74 sqsgycrnllynflkpkatrsincsgvtgrdgqavlaqalinnlevkkr-epftdhtys 132  
 QY 68 MTSQCDIYQTLRGYAOQLVSKSEKSPFIASLYVHKDAIWEELHAIYNQHNITVCIHYD 127  
 Db 133 ltrdcenhkaerfkldfpjskeevelfpaysmvlhexlenferllayvapqnlcyhvnd 192  
 QY 128 RKAPDTFKVAMNNLAKCFSNIFIAKSLAEVVAHISRLQADNCLSDLKSSIQMKRYVN 187  
 Db 193 eksptfkeavkaliscfpnvflasklvrvyvaswsvrgdlnmedllqssvwykfln 252  
 QY 188 LCGDPLKSNFELVSEIKLNGANMLETVPKPSKLERETVHHLELRVREYVKLPRT 247  
 Db 253 tctgdfipkksnaevqaklmngnsmesevppkhetrvkyhevr---dclhl---t 306  
 QY 248 NISKEAPPHNIQIFVGSAYFVLSQAFVYIFNNISIVODFFAMSKDYSPPDEHFWATLIRV 307  
 Db 307 nkkddppnylmtftgnayivasdfvghvklpkbsqgllewvxdtyspdehlwatlqra 366  
 QY 308 PGIREFI-SRGAQVSDLOSKTRLYKKNYEGFF-----YPSCTGSHLRSCVCTIGAAELR 361  
 Db 367 rmpgsvpnbpkkydidsmtslarlvkwsqhegdldkgapapcsqllnqralcvygsqdl 426  
 QY 362 WLKDGHWFAKFKDSKVDPIIKLAERL 390  
 Db 427 wmlqnhllankfdpkyddnalqcleeyl 455

RESULT 7  
 AAB18996  
 ID AAB18996 standard; Protein: 663 AA.  
 XX AAB18996;  
 AC AAB18996;  
 DF 08-FEB-2001 (first entry)  
 DE A partial core 2 beta-1,6-N-acetylglucosaminyltransferase.  
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLONAC-T;  
 KM cancer; cardiovascular disorder; inflammatory disorder; asthma;  
 KM rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;  
 KM septic shock; adult respiratory distress syndrome; ARDS; cancer;  
 KM platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;  
 KM clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;  
 KM diverticulitis; ulcerative colitis.  
 XX Homo sapiens.  
 OS CA2296936-A1.  
 PN 03-AUG-2000.  
 PD 03-FEB-2000; 2000CA-2236936.  
 PF 03-FEB-1999; 99US-0118674.  
 PR 03-FEB-1999; 99US-0118674.  
 XX

PA (GLYC-) GLYCODESIGN INC.

XX Korczak B, Lew A;

XX WPI: 2000-594746/57.

XX New nucleic acid molecules of core 2

XX beta-1,6-N-acetylglucosaminyltransferase useful for providing new

XX compositions for treatment of disorders mediated by the enzyme

XX including cancer, cardiovascular and inflammatory disorders.

XX Claim 3: Page 52; 66pp; English.

XX The present sequence represents a partial human core 2

XX beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.

XX The polypeptide can be used to treat diseases and disorders, such as

XX cancer, cardiovascular disorders and inflammatory disorders including

XX asthma, rheumatoid arthritis, inflammatory bowel disease,

XX arteriosclerosis, septic shock, adult respiratory distress syndrome

XX (ARDS) and cancer. Various platelet-mediated pathologies such as

XX atherosclerosis and clotting can also be treated. The polypeptides of

XX the invention are predominantly expressed in gastrointestinal tissue

XX (stomach, colon, intestine, testis) and are elevated in cancer.

XX Gastrointestinal disorders that may be prevented or treated include

XX ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis

XX and ulcerative colitis. The antibodies may be used in

XX immuno-histochemical analysis, to detect the novel polypeptide and to

XX localize it to particular cells and tissues and to specific subcellular

XX locations and to quantitate the level of expression.

XX Sequence 663 AA:

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Rat DH1 protein.

XX DH1: rat; screening; treatment; prevention; cardiomyopathy; inhibitor;

XX diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;

XX UDP-GlcNAc:galbeta1-3galNAc-alpha1; transgenic animal; germ line;

XX beta-1,6-N-acetylglucosaminyl-transferase.

XX Rattus sp.

XX CA2186987-A.

XX 02-APR-1998.

XX 02-OCT-1996; 96CA-2186987.

XX 02-OCT-1996; 96CA-2186987.

XX (MOUN ) MOUNT SINAI HOSPITAL CORP.

XX Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

XX WPI: 1998-399608/35.

XX N-PSDB; AAX24042.

XX Screening for substances that prevent or treat cardiomyopathy

XX associated with diabetes and hyperglycaemia - comprises reacting

XX core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor

XX in presence of test substance

XX Example 1; Fig 3A; 35pp; English.

XX This invention describes a method for screening for a substance that

XX may be used to prevent or treat cardiomyopathy associated with diabetes

XX and hyperglycaemia. This method involves reacting core 2 GlcNAc-T

XX (UDP-GlcNAc:galbeta1-3galNAc-alpha1; transgenic animal; germ line;

XX beta-1,6-N-acetylglucosaminyl-transferase) with an acceptor substrate and a sugar nucleotide donor in

XX the presence of a test substance under conditions whereby the core 2

XX GlcNAc-T produces a reaction product, determining the amount of reaction

XX product, and comparing the amount of reaction product with the amount

XX obtained in the absence of the test substance, where lower amounts of

XX reaction product in the presence of the test substance indicate that the

XX substance inhibits core 2 GlcNAc-T. The invention also describes (1)

XX methods for preventing or treating cardiomyopathy associated with

XX diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose

XX germ cells and somatic cells all contain a DNA construct introduced into

XX the animal or an ancestor of the animal at an embryonic stage, where

XX incorporation of the DNA construct into the germ line of the animal

XX causes the animal to develop cardiomyopathy similar to that associated

XX with diabetes mellitus and hyperglycaemia. This sequence represents the

XX rat DH1 protein which is used to describe the method of the invention.

XX Sequence 428 AA:

Query Match 38.9%; Score 854.5; DB 19; Length 428;  
Best Local Similarity 42.1%; Pred. No. 4.8e-69;  
Matches 177; Conservative 71; Mismatches 133; Indels 39; Gaps 10;

XX 1 POKDITLV-EYSLSTSPFVRNR-----YTHVK--DEVREYVNCSGIYEOEPLKIK- 48

XX 14 PKKYIYMTAVLVSILTSVTRHQKPEFVSVSHLSGDDPMNVCTKVLQGGPELQKV 73

XX 49 SLEI-----RRRDIILEDVVAVNTSPDCDIYQTLRGYAOQLVSKKEKSPFIANS 99

XX 74 KLEILTVGFKKPR-----CPHDYIMTRDCAAFIRTKYIMPELKEEGYFIAYSI 127

XX 100 VYHKDAIMVERLIHATYNOHNITCYHYDKRDPFKVAMNNLAKCFNSFIASKLEAVEY 159

XX 128 VVHKKIDMLDRLRALYMPGNYCIVHDKAEESFLAAGYIASCFDVFVSGLESVY 187

XX 160 AHISRLQADLNCISDLKSSIQMKYVINLGGDFPLKSNFELVSEIKINGANMLETYKP 219

RESULT 8

AAW93942

AAW93942 standard; Protein; 428 AA.

AAW93942;

28-JUN-1999 (first entry)

Db 188 aswsvxAdIncmkdlYrmanwkylinlcmgdfpiklnlelvrklskftgensletekm 247  
 QY 220 PMSKLEPFTYHHELRPRYEVYKLPRTNISKAPPHNIQIFVGSAYFLVLSQAFVYKIFN 279  
 Db 248 pmaeeewk-----krylvvdgkl-ttgyvkaqplktrpfssgafvvtreyvgvyle 301  
 QY 280 NSIVQDFEAFMSKDTYSPDEHEFNATLIRVPGIGEISRSQA-DVSDLOSKTRLYKWNYYEG 338  
 Db 302 nkniqkImewagdtyspdefiwtatqirpevgslpsshkydlstdmavarfvkwyfeg 361  
 QY 339 FF-----YPSCTGSHLSVCIYGAELRWLIDGHWFAKPFDSKVPDILIKLAETKEEQ 393  
 Db 362 dvsngapyppcgsvhvsvcvfygvdlsmlrkhhfrankfdmadvpfaigcleenlrhk 421  
 RESULT 9  
 AAM93943  
 ID AAM93943 standard; Protein: 428 AA.  
 XX  
 AC AAM93943;  
 XX  
 DT 28-JUN-1999 (first entry)  
 XX  
 DE Human core 2 Gnt protein.  
 XX  
 KW Screening; treatment; prevention; cardiomyopathy; inhibitor;  
 RW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;  
 KM UDP-GlcNAc:Galbeta1-3GalNAc-alpha1,6-N-acetylglucosaminyl-  
 beta-1,6-N-acetylglucosaminyl-transferase; human; core 2 Gnt.  
 XX  
 OS Homo sapiens.  
 XX  
 CA2186987-A.  
 XX  
 PD 02-APR-1998.  
 XX  
 PF 02-OCT-1996; 96CA-2186987.  
 XX  
 PR 02-OCT-1996; 96CA-2186987.  
 XX  
 PA (MOUN ) MOUNT SINAI HOSPITAL CORP.  
 XX  
 PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;  
 DR MPI: 1998-399608/35.  
 DR N-PSDB; AAX24043.  
 XX  
 PT Screening for substances that prevent or treat cardiomyopathy  
 associated with diabetes and hyperglycaemia - compounds reacting  
 with core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor  
 in presence of test substance  
 XX  
 PS Disclosure: Fig 9; 35pp; English.  
 XX  
 CC This invention describes a method for screening for a substance that  
 may be used to prevent or treat cardiomyopathy associated with diabetes  
 and hyperglycaemia. This method involves reacting core 2 GlcNAc-T  
 (UDP-GlcNAc:Galbeta1-3GalNAc-alpha1,6-N-acetylglucosaminyl-  
 transferase) with an acceptor substrate and a sugar nucleotide donor in  
 the presence of a test substance under conditions whereby the core 2  
 GlcNAc-T produces a reaction product, determining the amount of reaction  
 product, and comparing the amount of reaction product with the amount  
 obtained in the absence of the test substance, where lower amounts of  
 reaction product in the presence of the test substance indicate that the  
 substance inhibits core 2 GlcNAc-T. The invention also describes (1)  
 methods for preventing or treating cardiomyopathy associated with  
 diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose  
 germ cells and somatic cells all contain a DNA construct introduced into  
 the animal or an ancestor of the animal at an embryonic stage, where  
 incorporation of the DNA construct into the germ line of the animal  
 causes the animal to develop cardiomyopathy similar to that associated  
 with diabetes mellitus and hyperglycaemia. This sequence represents the  
 human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase)

CC which is used to describe the method of the invention.

XX Sequence 428 AA;

Query Match 38.7%; Score 849.5; DB 19; Length 428;  
 Best Local Similarity 43.0%; Pred. No. 1,4e-68;  
 Matches 181; Conservative 70; Mismatches 129; Indels 41; Gaps 11.

QY 1 POKDIYV-EYSLSTSPFVFNRYTHVKDE---VRY-----EVNCSGIYQEPLEI 46  
 Db 14 ptkyymfmlvlsistfsylr---lbgkpefsvvrhlslagengssdinctkylgqdvnei 70  
 QY 47 GK-----SEIFRRIIDLEDDDDVYAMSDCDIYOTLNGYAKQIYVSKESKSPPIASV 100  
 Db 71 qkvkllvllvkkfkrp--twtpddylnmsdsstikrkylyvepskseeafplaysiv 128  
 QY 101 VHKDAIMVERLHAIIYQNHIIYCIHYDRKAPDTFFKVMNNLAKCSNPIFASKLAEVEXA 160  
 Db 129 vhhklemldrlrlaympgnfyvvhvdtksedsyiaavmgiascfsvfvasrlsevyva 188  
 QY 161 HISRLQADNLCLSDLEKSSIQMKYVINLCQDPPKSNFELVSELKKNLGMAMLETVKRP 220  
 Db 189 swarvqadlncmkdlYamsanwkylinlcmgdfpiklnlelvrkllngennletermp 248  
 QY 221 NSKLEPFTYHHELRPRYEVYK-LPRTNISKAPPHNIQIFVGSAYFLVLSQAFVYKIF 278  
 Db 249 shheerw-----kryevvngkl-tngvtvmlpplleplsgaylvsvreygvyl 300  
 QY 279 NSIVQDFEAFMSKDTYSPDEHEFNATLIRVPGIGEISRSQA-DVSDLOSKTRLYKWNYYE 337  
 Db 301 qneikqlmewagdtyspdeylwatlqirpevgslpsshkydlstdmavarfvkwyfeg 360  
 QY 338 GFF-----YPSCTGSHLSVCIYGAELRWLIDGHWFAKPFDSKVPDILIKLAETKEE 392  
 Db 361 gdwskgapyppcdgvhvsvcvfygvdlnsmllrkhhfrankfdmadvpfaigcleenlrhk 420  
 QY 393 Q 393  
 Db 421 k 421  
 RESULT 10  
 AAB30298  
 ID AAB30298 standard; Protein: 428 AA.  
 XX  
 AC AAB30298;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 DE Human heart core 2 GlcNAc-T.  
 XX  
 KW Human; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;  
 RW UDP-GlcNAc:Galbeta1-3GalNAc-alpha1,6-N-acetylglucosaminyl-  
 N-acetylglucosaminyltransferase.  
 XX  
 OS Homo sapiens.  
 XX  
 US6131578-A.  
 XX  
 PD 17-OCT-2000.  
 XX  
 PF 02-OCT-1997; 97US-0943058.  
 XX  
 PR 02-OCT-1996; 96US-0046876.  
 XX  
 PA (KING/) KING G L.  
 PA (NISHU/) NISHIO Y.  
 PA (KOYA/) KOYA D.  
 PA (DENNIS/) DENNIS J W.  
 PA (WARREN/) WARREN C E.  
 XX  
 PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;

XX WPI: 2000-678642/66.  
 DR N-PSDB: AAC65469.  
 XX  
 PT Preventing or treating cardiomyopathy associated with diabetes mellitus  
 PT and hyperglycemia, comprises administering a substance that inhibits  
 PT core 2 N-acetylglucosamine-T activity -  
 PS  
 PS Claim 3: Fig 9: 21pp: English.  
 CC The present invention describes the human UDP-glucNAc:galbeta-1-3galNAc  
 CC alpha beta-6-N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein  
 CC and coding sequences. The enzyme is associated with cardiomyopathy in  
 CC diabetes and hyperglycaemia sufferers. The invention also provides  
 CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used  
 CC in treatment.  
 CC  
 XX  
 XX Sequence 428 AA:  
 SO  
 Query Match 38.7%; Score 849.5; DB 21; Length 428;  
 Best Local Similarity 43.0%; Pred. No. 1.4e-68;  
 Matches 181; Conservative 70; Mismatches 129; Indels 41; Gaps 11;  
 QY 1 POKDIYLV-EYSISTSPFVNRNRYTHWKDE---VRY-----EVNCSGIYCEPLEI 46  
 DB 14 ptkytmvavlsaltfsvlr---lhqkpefsvrnhlelagempssdlnctkylgqdvnei 70  
 QY 47 GK-----SLEIRRRDIILEDVVAMTSDCDIYOTLRGYAKLVSKKEKSPFIAYSLV 100  
 DB 71 qkvlelltvkfkfkrp--twtpddymntsdcsfkkrrylyveplskaeefplaysiv 128  
 QY 101 VHKDAIWEELHAIYQNHNIYCIHDKRAPDTEFKYAMNNLAKCFSTFIASKLEAVEYA 160  
 DB 129 vnhkiemldrlisaiyipnfyvhtdksedsyaaavgiascsnfvvaarlesvuya 188  
 QY 161 HISRLQADNLCLSDLKSSIQWKYVNLGCGDFPLKSNFELVSELKLNANMLFTVKNP 220  
 DB 189 swsrvgadlnckmdlyamsanwkyllnlgcmdfpikntlelvrkiklimgennletermp 248  
 QY 221 NSKIERETYNHELRLRVYEVY--KLPIRTNISKAPPHNIQIFVGSAYFLVLSQAEVKYIF 278  
 DB 249 shkeerw-----kkyevvngkl-tngtvtmklpplctplfsgsdyfvsreygyvl 300  
 QY 279 NNSIVODFFAMSKDYSPDEHFMTLLRVPGIPEISRSAQ-DVSDLOSKTRLVKNMYIE 337  
 DB 301 qnekiklmewagdytspdeylwatigripevpslpashkydlisdmgavarfvkwyfe 360  
 QY 338 GFF-----YPSCTGSHLRSCVIGAAELRWLIDGHWFAFKPDSKVPILIKCLAEKLEE 392  
 DB 361 gdvskgapypcdgwhvsvclfgagdlmmltkhhlfnakfdvadvdlfaigcldehlnh 420  
 QY 393 Q 393  
 DB 421 k 421  
 RESULT 11  
 AARS1386  
 ID AARS1386 standard; Protein: 428 AA.  
 AC AARS1386;  
 XX  
 XX 07-OCT-1994 (first entry)  
 DE Sequence of human core 2 beta 1-6  
 DE N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).  
 DE  
 DE C2GNT; 1-6 AGT; core 2 beta 1-6 N-acetylglucosaminyltransferase;  
 KM O-glycan.  
 XX  
 XX Homo sapiens.  
 OS  
 XX

EH Key Location/Qualifiers  
 FT Domain 10..32  
 FT /label= signal/membrane anchoring domain  
 FT Modified-site 52  
 FT /label= potential N-glycosylation site  
 FT /note= "see also Aas 58 and 95"  
 XX  
 XX EP590747-A.  
 XX  
 XX 06-APR-1994.  
 XX  
 XX 29-SEP-1993; 93EP-0250268.  
 XX  
 XX 01-OCT-1992; 92US-0955041.  
 XX  
 XX (JOL- ) LA JOLLA CANCER RES FOUND.  
 XX  
 XX Bierhuizen MFA, Fukuda M;  
 XX  
 XX WPI: 1994-111195/14.  
 DR N-PSDB: AAC61559.  
 XX  
 XX New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are  
 PT used for the study of the effect of variant O-glycan(s) on  
 PT cell-cell interactions, partic. in cancers  
 PS  
 PS Disclosure: Page 20-22; 34pp: English.  
 CC C2GNT, or an active fragment thereof, catalyses the formation of  
 CC critical branches in O-glycans. CDNA's encoding various  
 CC glycosyltransferases can be isolated by transient expression of cDNA  
 CC in recipient cells, e.g. COS-1. COS-1 cells were transfected  
 CC with a cDNA library, pCDSR alpha-2F1, constructed from poly(A)+ RNA  
 CC of activated T lymphocytes which express the C2GNT. Transfected  
 CC cells were selected using Mab T305, which identifies a  
 CC hexasaccharide on leukostatin. Leukostatin CD43 is an acceptor  
 CC molecule for C2GNT activity. A plasmid, pCD5R alpha-leu,  
 CC which directed expression of the T305 antigen was identified. The  
 CC cDNA insert was isolated and sequenced. The cDNA encoded the  
 CC acceptor molecule leukostatin CD43.  
 CC  
 XX  
 XX Sequence 428 AA:  
 SO  
 Query Match 38.3%; Score 840.5; DB 15; Length 428;  
 Best Local Similarity 42.8%; Pred. No. 8.8e-68;  
 Matches 180; Conservative 70; Mismatches 130; Indels 41; Gaps 11;  
 QY 1 POKDIYLV-EYSISTSPFVNRNRYTHWKDE---VRY-----EVNCSGIYCEPLEI 46  
 DB 14 ptkytmvavlsaltfsvlr---lhqkpefsvrnhlelagempssdlnctkylgqdvnei 70  
 QY 47 GK-----SLEIRRRDIILEDVVAMTSDCDIYOTLRGYAKLVSKKEKSPFIAYSLV 100  
 DB 71 qkvlelltvkfkfkrp--twtpddymntsdcsfkkrrylyveplskaeefplaysiv 128  
 QY 101 VHKDAIWEELHAIYQNHNIYCIHDKRAPDTEFKYAMNNLAKCFSTFIASKLEAVEYA 160  
 DB 129 vnhkiemldrlisaiyipnfyvhtdksedsyaaavgiascsnfvvaarlesvuya 188  
 QY 161 HISRLQADNLCLSDLKSSIQWKYVNLGCGDFPLKSNFELVSELKLNANMLFTVKNP 220  
 DB 189 swsrvgadlnckmdlyamsanwkyllnlgcmdfpikntlelvrkiklimgennletermp 248  
 QY 221 NSKIERETYNHELRLRVYEVY--KLPIRTNISKAPPHNIQIFVGSAYFLVLSQAEVKYIF 278  
 DB 249 shkeerw-----kkyevvngkl-tngtvtmklpplctplfsgsdyfvsreygyvl 300  
 QY 279 NNSIVODFFAMSKDYSPDEHFMTLLRVPGIPEISRSAQ-DVSDLOSKTRLVKNMYIE 337  
 DB 301 qnekiklmewagdytspdeylwatigripevpslpashkydlisdmgavarfvkwyfe 360  
 QY 338 GFF-----YPSCTGSHLRSCVIGAAELRWLIDGHWFAFKPDSKVPILIKCLAEKLEE 392

Db 361 gatskappypcdyghvsvclifegadlmwlrknhlfavdvjlfatqcldehnrh 420  
 QY 393 Q 393  
 Db 421 k 421

## RESULT 12

AAB30297  
 ID AAB30297 standard; Protein; 427 AA.

AC AAB30297;

DT 12-FEB-2001 (first entry)

DE Diabetic rat heart core 2 GlcNAc-T.

KW Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;  
 KM UDP-GlcNAc:Galbeta1-3GalNAcalpha1phar beta1-6  
 N-acetylglucosaminyltransferase.

OS Rattus sp.

PN US6131578-A.

PD 17-OCT-2000.

PF 02-OCT-1997; 97US-0943058.

PR 02-OCT-1996; 96US-0046876.

PA (KING/) KING G L.

PA (NISH/) NISHIO Y.

PA (KORA/) KORA D.

PA (DENN/) DENNIS J W.

PA (WARR/) WARREN C E.

PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;

DR WPI: 2000-678642/66.

DR N-PSDB: AAC65468.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 38.2%; Score 838; DB 21; Length 427;

Best Local Similarity 40.8%; Pred. NO. 1.5e-67;

Matches 172; Conservative 76; Mismatches 130; Indels 44; Gaps 11;

QY 1 POKDIIV-EYSLSTSPFNR-----YTHVK---DEVREVNCSGIYQEPLEIGKS 49

Db 14 pckyytmvvlsltftevrlnqkpefsvshleibgdqdsnnctkvgdpeelqkv 73

QY 50 LRRRDITLDD-----DDVYAMSDCDIYOTLGYKQKLVSKKESFPPIASLV 101

Db 74 ---Kxelltygekprpttphetumlrtdcasiflrkyimeptkeeygfplasyiv 129

QY 102 HKDAIVERLHAIVNOHNIYCIHYDRKAPDTFKVAMNNLAKCSNIPFASKLAVEVAH 161

Db 130 hkhidmldtllalymptqfycihvdtrkaeeeflaavvgiascfduvfvaqlesvvas 189  
 QY 162 ISRLQADNLCLSDLLKSSIQMKRYVINLCQDPPLKSNFELVSEKTLGAMMETVX-PP 220  
 Db 190 wsvvkdlnomkdyymannkylinlgcmfipkltlnleivrk1fsftgensletekmp 249  
 QY 221 NSKLERFTYHHELRVPEYV---KLPIRTNYSKAPPHNIQIFVGSAYEVLQAFVKY 277  
 Db 250 nke-----krwkyzyvvdgkl-tnlyvkaqpplktpilfsasayfvttteyygv 298  
 QY 278 FNNISIVQDFPANSKDTYSPDEHFMTLIRVPEIGELISRSQ-DVSDLSQKTRLVKMY 336  
 Db 299 lenkniqkimewagdtyspdeftwalqirdevp9slpsnhkyldsdmnavarfxwqif 358  
 QY 337 EGPF-----YPSCTGSHLRVSVCIYGAELRMLIKDGEHFAKFPDSKVPILIKLAEKLE 391  
 Db 359 egdvsngappccsgyghvsvclifegadlmwlrknhlfavdvjlfatqcldehnrh 418  
 QY 392 EQ 393  
 Db 419 hk 420

## RESULT 13

AAR71932  
 ID AAR71932 standard; Protein; 400 AA.

AC AAR71932;

DT 23-SEP-1995 (first entry)

DE I-branching enzyme.

KW Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGRT;

KW Inflammation; antiinflammatory; tumor; hypersensitivity; anemia;

KW transgenic animal; I antigen.

OS Homo sapiens.

PN WO9507020-A.

PD 16-MAR-1995.

PF 09-SEP-1993; 93WO-US08476.

PR 09-SEP-1993; 93WO-US08476.

PA (LJOL-) LA JOLLA CANCER RES FOUND.

PI Bierhuizen MFA, Fukuda M;

DR WPI: 1995-123182/16.

DR N-PSDB: AA089201.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Disclousure: Fig.6: 56pp; English.

CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells, was

CC reverse-transcribed and inserted into expression vector pCDNA1.

CC Plasmid DNA was used to transfect CHO-Py-leu cells. Transfected

CC cells were screened using human anti-I antigen antibodies and goat

CC anti-human IgM. Plasmid DNA was analysed to obtain DNA (given in

CC AA089201) encoding I-branching enzyme (AAR71932).

XX

Sequence 400 AA;

Query Match 32.4%; Score 711; DB 16; Length 400;

Best Local Similarity 43.7%; Pred. NO. 4.2e-56;

Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;



XX Blerhuizen MFA, Fukuda M;  
 XX WPI: 1998-361697/31.  
 DR N-PSDB; AAY30006.  
 XX

PT New nucleic acid sequences and their complementary sequences -  
 PT useful for producing fragment of recombinant human I-branching  
 PT -1,6-N-acetyl-glucosaminyl-transferase polypeptide  
 XX

PS Claim 14; Columns 31-34; 31p; English.

XX The present sequence represents human  
 CC beta-1,6-N-acetylglucosaminyltransferase, the I-branching enzyme (IGT).  
 CC The protein sequence has a type II transmembrane topology that consists  
 CC of a short amino terminal cytoplasmic sequence, a signal-anchor sequence  
 CC followed by a short stem region and a large carboxyl terminal catalytic  
 CC domain. The protein shows some homology to core  
 CC 2-beta-1,1-N-acetylglucosaminyltransferase (CGT). The nucleic acid  
 CC sequence is used for producing IGT polypeptides, which can be used  
 CC to treat disorders arising from under expression of IGT, e.g. increased  
 CC susceptibility to type-II hypersensitivity reactions such as neonatal  
 CC haemolytic disease, autoimmune haemolytic anaemia and thrombocytopenia.  
 XX

SQ Sequence 400 AA;

Query Match 32.4%; Score 711; DB 19; Length 400;

Best Local Similarity 43.7%; Pred. NO. 4.2e-56;

Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 70 SDCDIYQTLRGYAKLVSKESKSPFIAYSLVYHKDAIMYERLIHAIYNOHNYCIHYDRK 129  
 DB 70 sdcckyltqshytlaplskeeadiflayimvlnhndftfarllfaiympqnlvcvhdex 129  
 QY 130 APDTFKVAMNNLAKCFNSNFIASKLEAVEYAHISRLADLNCISDLKSSIQWKYVINC 189  
 DB 130 attefkaveqliscfnafaslakmepvyygslrqadlnclrdlsafeswkyvlnlc 189  
 QY 190 GQDFPLKSNPELVSELKUNGAMLETVPFNSKLERFTY-HHELRRVPEYVKLPRTN 248  
 DB 190 ggdipikltnkelvqykfgknlcpqvlppahalgtrkyvqenlgkelsyv---iflt 246  
 QY 249 ISKEAPPHNIQIFVGSAYEVLISQAFVKYIFNNSIVQDFPWSKDYSPDEHFMATLIRVP 308  
 DB 247 alkprrphnltlyfsgayalsrefanflvhdpravdlqwekdltfspdethwclnrlp 306  
 QY 309 GIGGEISRSKQDVSDLOSKTRLYKWNYYEGFFPSCGSHLSVCILYGAELRWLIKDG 368  
 DB 307 gvpqsmnpnas---wtgnlralksdmed-rhgchgnyvnglciyngdtkwlvnsps 360  
 QY 369 WFANKEDSKVPIILKCLAEKLEDOOR 395  
 DB 361 lfankfelntlyplvecl--elrher 385

Search completed: August 2, 2001, 14:44:51  
 Job time: 122 sec





```

XX  W0200114535-A2.
XX
XX  01-MAR-2001.
XX
XX  24-AUG-2000; 2000WO-DK00469.
XX
XX  24-AUG-1999; 99US-0150488.
XX
XX  (SCHW/) SCHWIENTEK T.
XX  (CLAU/) CLAUSEN H.
XX
XX  Schwientek T, Clausen H;
XX
XX  WPI: 2001-226615/23.
XX  N-PSDB; AAS00045.
XX
XX  New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
XX  useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
XX  thymus-related disorders, cancers, tumours, immunosuppression
XX
XX  Claim 17; Fig 1; 97pp; English.
XX
XX  The sequence represents Human UDP-N-acetyl-glucosamine:Galactose-beta1,
XX  3-N-acetylglucosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
XX  (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
XX  C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
XX  preparation of compositions for treating a conditions mediated by C2GnT3,
XX  particularly a thymus-related disorder. C2GnT3, nucleic acids
XX  encoding it and antibodies against it may also be used for in vitro
XX  purposes related to scientific research, DNA synthesis and manufacture of
XX  vectors, in the prognostic and diagnostic evaluation of conditions
XX  associated with altered expression or activity of C2GnT3 or conditions
XX  requiring modulation of C2GnT3, as well as in monitoring conditions by
XX  detecting and localising the DNA and protein. Disorders such as tumours
XX  (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
XX  breast or cervix), hypoactivity, hyperactivity, atrophy, enlargement of
XX  thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
XX  acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
XX  sepsis, wound healing, acute and chronic infection, cell-mediated or
XX  humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
XX  or nucleic acid. The antibodies may be used to screen potential
XX  therapeutic compounds to determine their effects on a conditions such as
XX  thymus-related disorder or cancer, to determine the level of C2GnT3
XX  expression in cells genetically engineered to produce C2GnT3, or to
XX  detect and quantify polypeptides in a sample to determine their role in a
XX  particular cellular events or pathological states and to diagnose and
XX  treat such pathological states.
XX
XX  Sequence 453 AA;
XX
XX  Query Match 100.0%; Score 2389; DB 22; Length 453;
XX  Best Local Similarity 100.0%; Pred. No. 1.6e-207;
XX  Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB  241 selklingammletvkpnskierftlyhbelrrpyevkrlpitnlskeepnqilfv 300
    |||
DB  301 GSAYFVLQAPFYKTFENNSIVQDFPANSKDTYSPDEHFWALLIVPGIIGCITSADVS 360
    |||
DB  301 GSAYFVLQAPFYKTFENNSIVQDFPANSKDTYSPDEHFWALLIVPGIIGCITSADVS 360
    |||
DB  361 DIQSKTRILVKNWYEGFYPSCGSHLSVCITYGAELRWLIKGDHWFANFDSKVPIL 420
    |||
DB  361 dIqsktrilvknwyegfypscgshlsrvcilygaaelrwlkdgwhfankfcdskvpil 420
    |||
DB  421 IKCLAEKLEEQQDQWITLPSEKLEMDRNLTTS 453
    |||
DB  421 IKCLAEKLEEQQDQWITLPSEKLEMDRNLTTS 453
    |||

RESULT 2
AAB18995
ID  AAB18995 standard; Protein: 438 AA.
XX
XX  AAB18995;
XX
XX  08-FEB-2001 (first entry)
XX
XX  A core 2 beta-1,6-N-acetylglucosaminyltransferase.
XX
XX  Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
XX  cancer; cardiovascular disorder; inflammatory disorder; asthma;
XX  rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
XX  septic shock; adult respiratory distress syndrome; AIDS; cancer;
XX  platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
XX  clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
XX  diverticulitis; ulcerative colitis.
XX
XX  Homo sapiens.
XX
XX  CA2296936-A1.
XX
XX  03-AUG-2000.
XX
XX  03-FEB-2000; 2000CA-2296936.
XX
XX  03-FEB-1999; 99US-0118674.
XX
XX  (GLYC-) GLYCDESIGN INC.
XX
XX  Korczak B, Lew A;
XX
XX  WPI: 2000-594746/57.
XX  N-PSDB; AAA96569.
XX
XX  New nucleic acid molecules of core 2
XX  beta-1,6-N-acetylglucosaminyltransferase useful for providing new
XX  compositions for treatment of disorders mediated by the enzyme
XX  including cancer, cardiovascular and inflammatory disorders.
XX
XX  Claim 3; Page 50-51; 66pp; English.
XX
XX  The present sequence represents a human core 2
XX  beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
XX  The polypeptide can be used to treat diseases and disorders, such as
XX  cancer, cardiovascular disorders and inflammatory disorders including
XX  asthma, rheumatoid arthritis, inflammatory bowel disease,
XX  arteriosclerosis, septic shock, adult respiratory distress syndrome
XX  (ARDS) and cancer. Various platelet-mediated pathologies such as
XX  atherosclerosis and clotting can also be treated. The polypeptides of
XX  the invention are predominantly expressed in gastrointestinal tissue
XX  (stomach, colon, intestine, testis) and are elevated in cancer.
XX  gastrointestinal disorders that may be prevented or treated include
XX  ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
XX  and ulcerative colitis. The antibodies may be used in
XX  immuno-histochemical analysis, to detect the novel polypeptide and to
XX  localize it to particular cells and tissues and to specific subcellular

```





Query Match 36.3%; Score 868; DB 21; Length 663;  
 Best Local Similarity 43.0%; Pred. No. 8.1e-70;  
 Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

Y 23 LMLSLKLKLVN-----RRLEPKDIDYLVYSLSTSPFVRN-RYTHVKNDEVREVCSCG 75  
 Db 97 lwaigcmllatvalklstfrlckdsdhlglstresgqycnlllynfikltpakrsincsg 156  
 Y 76 IY--EOEPL--EIGKSEIRRRDIIDDDDVAMTSDCDIYQTLRGYAKLVSKESKSF 131  
 Db 157 vtrgdeevlqalinnlevkkkr-epfcdthylstltdcehfkaerffiqfplskeevf 215  
 Y 132 FIATSLVYHKAIMVERLIHAIYQNHNYICHHYDRKAPDTFKVAMNNLACFSNFIASK 191  
 Db 216 playsmvtheklenferlliravapqnlcyvhvdekspetkkeavkaliscfpvltask 275  
 Y 192 LEAVEYAHISRLOADLNCISDLKSSIQMKYVINLGGDFPLKSNFELVSELRKLGAMM 251  
 Db 276 lrvvyaswstrvgadlnmedllgsvpmkyflntcgdfdkanaemvgalkmlngns 335  
 Y 252 LETVAPPNSKLERFTYHHELRVPEYVKLPRTINISKEAPPNIQIFVGSATVLSQAF 311  
 Db 336 mesevppkhketrkwykhevr--dlhl--lnkkkdpplnltmftgnaylvaaardf 389  
 Y 312 VKYIFNNSIVODEFAMSKDTYSPDEHFWATLIRVGPGEI-SRSADVDSDLOSKTRLVK 370  
 Db 390 vghvknkpsqgllewvkdtyspdehfwatlgrrampgsvpnhpkysdmtslarlvk 449  
 Y 371 WNYVGEF-----YPSCTGSHLSVCIYGAELRWLKDGMFWANKFDSKVDPLIKCLA 425  
 Db 450 wqghgedldkqapypcsglbrqalcvygagdlmwllqnhlhlnkfkpkyvdnaqlcle 509  
 Y 426 EKL 428  
 Db 510 eyl 512

RESULT 6  
 AAM93942  
 ID AAM93942 standard; Protein; 428 AA.

AC AAM93942;  
 DT 28-JUN-1999 (first entry)  
 XX  
 DE Rat DH1 protein.  
 KM DH1; rat; screening; treatment; prevention; cardiomyopathy; inhibitor;  
 KM diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;  
 KM UDP-GlcNAc:Galbeta1-3GalNAc-alphaR; transgenic animal; germ line;  
 KM beta-1,6-N-acetylglucosaminyl-transferase.  
 OS Rattus sp.  
 XX  
 CA2186987-A.  
 PN  
 XX  
 PD 02-APR-1998.  
 PF 02-OCT-1996; 96CA-2186987.  
 PR 02-OCT-1996; 96CA-2186987.  
 XX  
 PA (MOUN ) MOUNT SINAI HOSPITAL CORP.  
 PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;  
 DR WPI; 1998-399608/35.  
 DR N-PSDB; AAX24042.  
 XX  
 PT Screening for substances that prevent or treat cardiomyopathy  
 associated with diabetes and hyperglycaemia - comprises reacting  
 core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor  
 PT

PT in presence of test substance  
 XX  
 PS Example 1; Fig 3A; 35pp; English.  
 XX  
 CC This invention describes a method for screening for a substance that  
 CC may be used to prevent or treat cardiomyopathy associated with diabetes  
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T  
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alphaR beta-1,6-N-acetylglucosaminyl-  
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in  
 CC the presence of a test substance under conditions whereby the core 2  
 CC GlcNAc-T produces a reaction product, determining the amount of reaction  
 CC product, and comparing the amount of reaction product with the amount  
 CC obtained in the absence of the test substance, where lower amounts of  
 CC reaction product in the presence of the test substance indicate that the  
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)  
 CC methods for preventing or treating cardiomyopathy associated with  
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose  
 CC germ cells and somatic cells all contain a DNA construct introduced into  
 CC the animal or an ancestor of the animal at an embryonic stage, where  
 CC incorporation of the DNA construct into the germ line of the animal  
 CC causes the animal to develop cardiomyopathy similar to that associated  
 CC with diabetes mellitus and hyperglycaemia. This sequence represents the  
 CC rat DH1 protein which is used to describe the method of the invention.  
 CC  
 XX  
 SQ Sequence 428 AA;

Query Match 36.3%; Score 866.5; DB 19; Length 428;  
 Best Local Similarity 42.3%; Pred. No. 5.8e-70;  
 Matches 183; Conservative 73; Mismatches 136; Indels 41; Gaps 11;

Y 28 LKLLNVRRLF--PQKIDIVLV-EYSLSTSEFVRN-----YTHVK---DEVRYVNCSS 74  
 Db 1 mlnlfrlrlfsytkyfmvlvlsiltfsvrlnhqkpefvsyshleagddpnsnvt 60  
 Y 75 GIYQGEPLLEIGK-SLEI-----RRRIIDLEDDDVAMTSDCDIYQTLRGYAKLV 124  
 Db 61 kvlggdpelqkvklellvtqfkkrr-----tphdyimrrdcasfirttkymlepl 114  
 Y 125 SKEEKSPPIAYSLVHNDAMIVERLIHAIYQNHNYICHHYDRKAPDTFKVAMNNLAKCS 184  
 Db 115 tkeevgfpilayslvvhhkldmlrlalypmqpfycihvdrkaesflaavvgiascd 174  
 Y 185 NIFIAKLEAVEYAHISRLOADLNCISDLKSSIQMKYVINLGGDFPLKSNFELVSELR 244  
 Db 175 nvfvaqslsvvyaswstrvgadlnmedllgsvpmkyflntcgdfdkanaemvgalkmlngns 234  
 Y 245 KLGANMLETVKPPNSKLERFTYHHELRVPEYVKLPRTINISKEAPPNIQIFVGSAY 304  
 Db 235 sfgensletemkmpnkeerwk-----krytvtvdgkl-tlgyvkaqplktdlfsagay 288  
 Y 305 FYLSQAIVKILFNNSIVODEFAMSKDTYSPDEHFWATLIRVGPGEISRSAG-DVSDIQ 363  
 Db 289 fvtvltreyvgyvlenknlqkfmewagdcyspdehfwatlgripvpslpskhydlslm 348  
 Y 364 SKTRLVKNYVGEF-----YPSCTGSHLSVCIYGAELRWLKDGMFWANKFDSKVD 418  
 Db 349 avarfvkwyfegdvngapypcsglbrqalcvygagdlmwllqnhlhlnkfkpkyvdnaqlcle 408  
 Y 419 ILIKCLAEKLEEQ 431  
 Db 409 falqlceehlrhk 421

RESULT 7  
 AAB54344  
 ID AAB54344 standard; Protein; 465 AA.

AC AAB54344;  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.  
 DE

XX	Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW	detection; diagnosis; identification; cytostatic; neuroprotective;
KW	neotrophic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW	antiinflammatory; cardiac; gene therapy; chromosome mapping;
KW	linkage analysis; tissue identification; tissue typing; forensic;
KW	neural; immune system; muscular; reproductive; gastrointestinal;
KW	pulmonary; cardiovascular; renal; proliferative.
XX	
OS	Homo sapiens.
XX	
PN	WO20005520-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US05989.
XX	
PR	12-MAR-1999; 99US-0124270.
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-579444/54.
DR	N-PSDB; AAC99109.
XX	
PT	New nucleic acid that is a pancreatic cancer antigen for preventing,
PT	treating, or ameliorating a medical condition, particular pancreatic
PT	cancer, or for use in assays for diagnosing a pathological condition
XX	
SI	Claim 11; Page 1237-1239; 1379pp; English.

OY	192	LEAVEAHISLOADDLNCILSLKSSLSIOWKVINLCODEPKSNFELVSEELKLNGANK	25
Dd	219	Ivrvyaaawstrvgddlnomedllqssrpbwkyflntcgdtfpiksnmaeygalkmlngns	278
OY	232	LEIVAKPPNSKIERTTYHHELFRRVREVKLPFRINISKEAPPHNIQIFVSASAFVLSOAF	311
Dd	279	meseppbhkietrkynhevr---dlhlh--fnkkdkpppyltnftgnaylvaseidf	322
OY	312	VKYIRNNISIVODPFAMSKDRTYSPODEFMATLRVPGLPGEI-SRSDOVGDLOSCTRVLX	370
Dd	333	vghvhtknpsqgllewvkdytspehnkatgrtgmppsrgmnbphkydisdmcsiarlvk	392
OY	371	WNYYZGF-----YPCTGSHRSVCVIYGAEFLWLIKDGHWFAFKRKDSKVDPLILCLA	425
Dd	393	wqghgedldxgapapocsgjnhraicvyagadlnmmiqnhiilankridrpvdnalgcle	452
OY	426	EKL 428	
Dd	453	eyl 455	

Query Match	36.28	Score 866	DB 21	Length 465
Best Local Similarity	43.04	Pred. No. 7.3e-70		
Matches 182	Conservative 74	Mismatches 143	Indels 24	Gaps 10
Sequence	465 AA			
23	EWLWS--LKLNLVRRLLFPQK--DIYLVYSLSISPEYRN-RYTHVAKDEVREYNSG	75		
DB	40 lwalgcmjllatvalaklsfxlkcdshlglesresgsgymnlyflfpakzinscg	99		
QY	76 IY--QOEP--EIGKSEIRRDIIEDDDVAMSDCIIYOTLTGVAQKIVSKSEKFS	131		
DB	100 vrrdgaeavlaqnlmnlvckkr-epftcdhyslstrdehfaekxfqfplskseerf	158		
QY	132 PAYSIVYHKAQIINVERTLIIATYQNHNIYCIYIHDRKAPDFTFKVAMNNLAKCSNIFLASK	191		
DB	159 playsmvtshetienferlliravypqnlcyhvdkspefkvaavallscfpvflask	218		

RESULT 8  
 ID AAM93943 standard; Protein; 428 AA.  
 AC AAM93943;  
 DT 28-JUN-1999 (first entry)  
 DE Human core 2 Gnt protein.  
 XX  
 XX Screening; treatment; prevention; cardiomyopathy; inhibitor;  
 KM diabetes mellitus; hyperglycemia; core 2 GlcNAc-T; acceptor substrate;  
 KM UDP-GlcNAc:Galbeta1-3GalNAc-alpha; transgenic animal; germ line;  
 KM beta-1,6-N-acetylglucosaminyl-transferase; human; core 2 Gnt.  
 OS  
 XX Homo sapiens.  
 XX CA2186987-A.  
 PN  
 XX 02-APR-1998.  
 XX  
 XX 02-OCT-1996; 96CA-2186987.  
 PF  
 PR 02-OCT-1996; 96CA-2186987.  
 XX  
 PA (MOUN ) MOUNT SINAI HOSPITAL CORP.  
 XX  
 PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;  
 DR WPI; 1998-399608/35.  
 XX N-PSDB; AAX24043.  
 XX  
 PT Screening for substances that prevent or treat cardiomyopathy  
 PT associated with diabetes and hyperglycemia - comprises reacting  
 PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor  
 PT in presence of test substance  
 PT  
 PS  
 PS Disclosure; Fig 9: 35pp; English.  
 XX  
 XX This invention describes a method for screening for a substance that  
 CC may be used to prevent or treat cardiomyopathy associated with diabetes  
 CC and hyperglycemia. This method involves reacting core 2 GlcNAc-T  
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alpha; beta-1,6-N-acetylglucosaminyl-  
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in  
 CC the presence of a test substance under conditions whereby the core 2  
 CC GlcNAc-T produces a reaction product, determining the amount of reaction  
 CC product, and comparing the amount of reaction product with the amount  
 CC obtained in the absence of the test substance, where lower amounts of  
 CC reaction product in the presence of the test substance indicate that the  
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)  
 CC methods for preventing or treating cardiomyopathy associated with

CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose germ cells and somatic cells all contain a DNA construct introduced into the animal or an ancestor of the animal at an embryonic stage, where incorporation of the DNA construct into the germ line of the animal causes the animal to develop cardiomyopathy similar to that associated with diabetes mellitus and hyperglycaemia. This sequence represents the human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase) which is used to describe the method of the invention.

XX Sequence 428 AA;

Query Match 36.1%; Score 863.5; DB 19; Length 428;  
Best Local Similarity 43.1%; Pred. No. 1,1e-69;  
Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

QY 28 LKLLINRLR--PQKDIYLV-EYSLSSTSPFVNRRTYHKDE---VRY-----EV 71  
DB 1 mlrtllrrrlfssyptkylfmlvlslltfsvlr---lhqkpefsvrhlagenpsd 57  
QY 72 NCSGIYEQEPLKIGK-----SLKRRRDIIDDDVYVMTSDCDIYQTLRGYAKLYS 125  
DB 58 nctkvlgdvneiqkvlelltvkfkrrp--rtppdyimtsccsfikrkyiveps 115  
QY 126 KEKSPFIAYSLVYHKDAIMVERLTHAIYNOHNYCIHYDRKAPDTEKVMNNLAKCFSN 185  
DB 116 keeaeffplaysivvnhklemldrllraltymqnfycvhtkksedyslaavmgiascfn 175  
QY 166 IFIAKLAIVAYAHISRLQADLNCSDILKSSIQWKYVINLCGQDFPLKSNFELVSELK 245  
DB 176 vfasrlesvvyasvsvgadlncmkdlyamsanwkyllnlgmdfpiknlsvtkl 235  
QY 246 LINGAMLETVKPPNSKLERFTYHHELRVPYEV--KLPIRTNISKEAPPNIIQIVGSA 303  
DB 236 lmgennletermphskeerw-----kryevvngkl-tntgtvkmplpletplfsgsa 287  
QY 304 YFVLSQAFVKYIFNNSTVQDFPFAWSKDTYSPDEHFWATLIRVPGIGELISRSAG-DVSD 362  
DB 288 yfvsvreyvgvylgnekqlmewagdtyspdeylwatigrilpevpaspashkylsdm 347  
QY 363 QSKTRLVKNWNYEGF-----YPSCTGSHLSVSCIYGAELRWLKDGMFANKFDSKVD 417  
DB 348 qavarfvkwgyfegdvskgapypcdgvhvrsvclfgaglnmllrkxhllfankfdvcd 407  
QY 418 PILIKCLAELKEEQ 431  
DB 408 lfaicldenhtrk 421

RESULT 9  
AAB30298  
ID AAB30298 standard; Protein; 428 AA.

AC AAB30298;  
XX 12-FEB-2001 (first entry)  
DT 12-FEB-2001 (first entry)  
XX Human heart core 2 GlcNAc-T.  
DE Human heart core 2 GlcNAc-T.  
XX Human; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;  
KM UDP-GlcNAc:galbetal-3galNacalpinar betal-6  
KW N-acetylglucosaminyltransferase.  
XX Homo sapiens.  
OS Homo sapiens.  
XX US6131578-A.  
PN 17-OCT-2000.  
PD 17-OCT-2000.  
XX 02-OCT-1997; 97US-0943058.  
PF 02-OCT-1997; 96US-0046876.  
XX 02-OCT-1996;  
PR 02-OCT-1996;  
XX

PA (KING/) KING G L.  
PA (NISH/) NISHIO Y.  
PA (KOYA/) KOYA D.  
PA (DENN/) DENNIS J W.  
PA (WARR/) WARREN C E.  
XX Nishio Y, Koya D, King GL, Warren CE, Dennis JW;  
PI WPI; 2000-678642/66.  
DR N-PSDB; AAC65469.  
XX

PT Preventing or treating cardiomyopathy associated with diabetes mellitus and hyperglycaemia, comprises administering a substance that inhibits core 2 N-acetylglucosamine-T activity -

PS Claim 3; Fig 9; 21pp; English.

CC The present invention describes the human UDP-GlcNAc:galbetal-3galNac alphaar betal-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein and coding sequences. The enzyme is associated with cardiomyopathy in diabetes and hyperglycaemia sufferers. The invention also provides methods for identifying inhibitors of core 2 GlcNAc-T which can be used in treatment.

XX Sequence 428 AA;

Query Match 36.1%; Score 863.5; DB 21; Length 428;  
Best Local Similarity 43.1%; Pred. No. 1,1e-69;  
Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

QY 28 LKLLINRLR--PQKDIYLV-EYSLSSTSPFVNRRTYHKDE---VRY-----EV 71  
DB 1 mlrtllrrrlfssyptkylfmlvlslltfsvlr---lhqkpefsvrhlagenpsd 57  
QY 72 NCSGIYEQEPLKIGK-----SLKRRRDIIDDDVYVMTSDCDIYQTLRGYAKLYS 125  
DB 58 nctkvlgdvneiqkvlelltvkfkrrp--rtppdyimtsccsfikrkyiveps 115  
QY 126 KEKSPFIAYSLVYHKDAIMVERLTHAIYNOHNYCIHYDRKAPDTEKVMNNLAKCFSN 185  
DB 116 keeaeffplaysivvnhklemldrllraltymqnfycvhtkksedyslaavmgiascfn 175  
QY 166 IFIAKLAIVAYAHISRLQADLNCSDILKSSIQWKYVINLCGQDFPLKSNFELVSELK 245  
DB 176 vfasrlesvvyasvsvgadlncmkdlyamsanwkyllnlgmdfpiknlsvtkl 235  
QY 246 LINGAMLETVKPPNSKLERFTYHHELRVPYEV--KLPIRTNISKEAPPNIIQIVGSA 303  
DB 236 lmgennletermphskeerw-----kryevvngkl-tntgtvkmplpletplfsgsa 287  
QY 304 YFVLSQAFVKYIFNNSTVQDFPFAWSKDTYSPDEHFWATLIRVPGIGELISRSAG-DVSD 362  
DB 288 yfvsvreyvgvylgnekqlmewagdtyspdeylwatigrilpevpaspashkylsdm 347  
QY 363 QSKTRLVKNWNYEGF-----YPSCTGSHLSVSCIYGAELRWLKDGMFANKFDSKVD 417  
DB 348 qavarfvkwgyfegdvskgapypcdgvhvrsvclfgaglnmllrkxhllfankfdvcd 407  
QY 418 PILIKCLAELKEEQ 431  
DB 408 lfaicldenhtrk 421

RESULT 10  
AAB18999  
ID AAB18999 standard; Peptide; 406 AA.

AC AAB18999;  
XX 08-FEB-2001 (first entry)  
DT 08-FEB-2001 (first entry)  
XX A core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region.  
DE

```

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis.
XX
OS Homo sapiens.
XX
CA2296936-A1.
XX
PD 03-AUG-2000.
XX
PF 03-FEB-2000; 200CCA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX
PA (GLYC-) GLYCODESIGN INC.
XX
PI Korczak B, Lew A;
XX
DR WPI; 2000-594746/57.
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
PS Claim 3; Page 53; 66pp; English.
XX
CC The present sequence is derived from a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 406 AA;

```

```

Db 248 nkkddpppyhlfttgnayivasrdtvgvhlknpksqqliewkdkyspdenhwatclqra 307
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 346 PGRIGEY-SRSACVPSLOSKTRILVKNMYEGFF-----YPSCTGSHRSVCYGAELR 399
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 308 rmpgsvpnphkpydsmtslarlvkxqghegdldkqapapsghlqratcyvgsgdln 367
OY 400 WLIDGHWPFANKPDSKYDPILIKCLAKL 428
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 368 wmlqnhllanktkpkyddnalgcleeyl 396
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 11.
ARS1386
ID ARS1386 standard; Protein; 428 AA.
XX
AC ARS1386;
XX
DT 07-OCT-1994 (first entry)
XX
DE Sequence of human core 2 beta 1-6
DE N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).
XX
KW C2GNT; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;
KW O-glycan.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Domain 10..32
FH Modified-site 52
FH FT /label= potential N-glycosylation site
FH FT /note= "see also AAs 58 and 95"
XX
PN EP590747-A.
XX
PD 06-APR-1994.
XX
PF 29-SEP-1993; 93EP-0250268.
XX
PR 01-OCT-1992; 92US-0955041.
XX
PA (JOL-) LA JOLLA CANCER RES FOUND.
XX
PI Biehuizen MFA, Fukuda M;
XX
DR WPI; 1994-111195/14.
XX
DR N-PSDB; AA061559.
XX
PT New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are
PT used for the study of the effect of variant O-glycan(s) on
PT cell-cell interactions, partic. in cancers
XX
PS Disclosure; Page 20-22; 34pp; English.
XX
CC C2GNT, or an active fragment thereof, catalyses the formation of
CC critical branches in O-glycans. cDNAs encoding various
CC glycosyltransferases can be isolated by transient expression of cDNA
CC in recipient cells, e.g. COS-1. COS-1 cells were transfected
CC with a cDNA library, pCDSR alpha-2F1, constructed from poly(A)+ RNA
CC of activated T lymphocytes which express the C2GNT. Transfected
CC cells were selected using MAb T305, which identifies a
CC hexasaccharide on leukosialin. Leukosialin CD43 is an acceptor
CC molecule for C2GNT activity. A plasmid, pCDSR alpha-leu,
CC which directed expression of the T305 antigen was identified. The
CC cDNA insert was isolated and sequenced. The cDNA encoded the
CC acceptor molecule leukosialin CD43.
XX
SQ Sequence 428 AA;

```

Query Match 35.8%; Score 854.5; DB 15; Length 428;





PR 09-SEP-1993: 93WO-US08476.  
 XX (LJOL-) LA JOLLA CANCER RES FOUND.  
 XX  
 PI Bierhuizen MFA, Fukuda M;  
 XX  
 DR WPI: 1995-123182/16.  
 DR N-PSDB; AA089201.  
 XX  
 PT Nucleic acid encoding I-branching enzyme - used to develop prods.  
 PT for treating e.g. inflammatory responses, tumours or  
 PT hypersensitivity reactions  
 XX  
 PS Disclosure; Fig.6: 56pp; English.  
 XX  
 CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells was  
 CC reverse-transcribed and inserted into expression vector pCDNA1.  
 CC Plasmid DNA was used to transfect CHO-Py-leu cells. Transfected  
 CC cells were screened using human anti-I antigen antibodies and goat  
 CC anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in  
 CC AA089201) encoding I-branching enzyme (AA071932).  
 CC  
 XX  
 SQ Sequence 400 AA;

Query Match 29.8%; Score 711; DB 16; Length 400;  
 Best Local Similarity 43.7%; Pred. No. 5.9e-56;  
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SDCDIYOTLNGYAOKLVSKKEKSPFIAYSLVVKDAIMVERLLHAIYNOHNYCIHYDRK 167  
 DB 70 ssckeyltqshytlaplskeeadfplaylmvlnhdtfarlfralypmqnlycvhvdex 129  
 QY 168 APDFFKAMNNLAKCFSENFIFASKLEAVEYAHISRLOADNCLSDLKSSIQKRYVINC 227  
 DB 130 atetfkaveqliscfnafaskmepvyggsrlqadlnclrdtsafevskxyintc 189  
 QY 228 GODEPLKSNFELVSELKNGANLLEVKPNSKLEFRTY-HHELRRVPEYVKLPRTN 286  
 DB 190 gddfpkltnkeivqlykgfkgknltpgvlpahalgtrkyvqehlgkelsy---lftt 246  
 QY 287 ISKEAPPHNIQIFVGSAYFLVSOAFVXYIFENNSIVODFFAMSKDTPSPDEHFMTLIRVP 346  
 DB 247 alkpppphnltyfgsayaalsteifanfyldpravdlqskdtfspehfwltnirp 306  
 QY 347 GIPGEISRSADVDSDQSKTRLVKWNYYEGFFYPSCTGSHRSVCITGAELRWLTKDGH 406  
 DB 307 gvpqsmnas-----wtgnlraikwsdmed-rhgqchgyhvngiclyngdtklwlvnsp 360  
 QY 407 WFANKFDSKYDPLIKICLAEKLEEQOR 433  
 DB 361 lfankfelntlyplvecl--elrhner 385

RESULT 14  
 AAR92474  
 ID AAR92474 standard; Protein: 400 AA.  
 XX  
 AC AAR92474;  
 XX  
 DT 22-APR-1996 (first entry)  
 XX  
 DE Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme.  
 XX  
 KW Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; Ignt;  
 XX glycosyltransferase; blood group; I antigen; poly(lactosaminoglycan).  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 7..25  
 FT /label= Membrane-anchoring\_domain  
 FT Modified-site 37

FT Modified-site 212  
 FT /label= N-glycosylation\_site  
 FT Modified-site 253  
 FT /label= N-glycosylation\_site  
 FT Modified-site 314  
 FT /label= N-glycosylation\_site  
 FT Modified-site 388  
 FT /label= N-glycosylation\_site  
 XX  
 FN US5484590-A.  
 XX  
 PD 16-JAN-1996.  
 XX  
 PF 09-SEP-1993; 93US-0118906.  
 XX  
 PR 09-SEP-1993; 93US-0118906.  
 XX  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 XX  
 PI Bierhuizen MFA, Fukuda M;  
 XX  
 DR WPI: 1996-087019/09.  
 DR N-PSDB; AAT16201.  
 XX  
 PT Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme  
 PT - used to develop prods. for the study, detection and treatment of  
 PT pathological conditions involving the enzyme.  
 XX  
 PS Claim 1; Column 31-34; 29pp; English.  
 XX  
 CC Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme  
 CC (Ignt) (AAR92474) is an enzyme that converts blood group I antigen to  
 CC I antigen. The occurrence of I antigen is associated with  
 CC development and maturation of erythroid cells. Membrane-bound or  
 CC soluble forms of Ignt are obtd. by expression of encoding DNA  
 CC sequences (see AAT16201) in host cells. Ignt is used to study the  
 CC role of Ignt in development and oncogenesis, to alleviate disorders  
 CC associated with Ignt underexpression (e.g. haemolytic disease of  
 CC the newborn, haemolytic anemia and thrombocytopenia), to regulate  
 CC Ignt expression, and in detection and diagnostic applications.  
 CC  
 XX  
 SQ Sequence 400 AA;

Query Match 29.8%; Score 711; DB 17; Length 400;  
 Best Local Similarity 43.7%; Pred. No. 5.9e-56;  
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SDCDIYOTLNGYAOKLVSKKEKSPFIAYSLVVKDAIMVERLLHAIYNOHNYCIHYDRK 167  
 DB 70 ssckeyltqshytlaplskeeadfplaylmvlnhdtfarlfralypmqnlycvhvdex 129  
 QY 168 APDFFKAMNNLAKCFSENFIFASKLEAVEYAHISRLOADNCLSDLKSSIQKRYVINC 227  
 DB 130 atetfkaveqliscfnafaskmepvyggsrlqadlnclrdtsafevskxyintc 189  
 QY 228 GODEPLKSNFELVSELKNGANLLEVKPNSKLEFRTY-HHELRRVPEYVKLPRTN 286  
 DB 190 gddfpkltnkeivqlykgfkgknltpgvlpahalgtrkyvqehlgkelsy---lftt 246  
 QY 287 ISKEAPPHNIQIFVGSAYFLVSOAFVXYIFENNSIVODFFAMSKDTPSPDEHFMTLIRVP 346  
 DB 247 alkpppphnltyfgsayaalsteifanfyldpravdlqskdtfspehfwltnirp 306  
 QY 347 GIPGEISRSADVDSDQSKTRLVKWNYYEGFFYPSCTGSHRSVCITGAELRWLTKDGH 406  
 DB 307 gvpqsmnas-----wtgnlraikwsdmed-rhgqchgyhvngiclyngdtklwlvnsp 360  
 QY 407 WFANKFDSKYDPLIKICLAEKLEEQOR 433  
 DB 361 lfankfelntlyplvecl--elrhner 385

RESULT 15

AAW56628 standard; Protein; 400 AA.

AAW56628;

11-AUG-1998 (first entry)

Full length amino acid sequence of human IGNT.

Human: beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGNT; core 2-beta-1,1-N-acetylglucosaminyltransferase; C2GNT; treatment; disorder; under expression; type-II hypersensitivity reaction; neonatal haemolytic disease; autoimmune haemolytic anaemia; thrombocytopenia.

Homo sapiens.

US5766910-A.

16-JUN-1998.

07-JUN-1995; 95US-0488135.

09-SEP-1993; 93US-0118906.

(LJOL-) LA JOLLA CANCER RES FOUND.

Bierhuizen MFA, Fukuda M;

MP1: 1998-361697/31.

N-PSDB: AAV30006.

New nucleic acid sequences and their complementary sequences - useful for producing fragment of recombinant human I-branching -1,6-N-acetyl-glucosaminyl-transferase polypeptide

Claim 14; Columns 31-34; 31pp; English.

The present sequence represents human beta-1,6-N-acetylglucosaminyltransferase, the I-branching enzyme (IGNT). The protein sequence has a type II transmembrane topology that consists of a short amino terminal cytoplasmic sequence, a signal anchor sequence followed by a short stem region and a large carboxyl terminal catalytic domain. The protein shows some homology to core 2-beta-1,1-N-acetylglucosaminyltransferase (C2GNT). The nucleic acid sequence is used for producing IGNT polypeptides, which can be used to treat disorders arising from under expression of IGNT, e.g. increased susceptibility to type-II hypersensitivity reactions such as neonatal haemolytic disease, autoimmune haemolytic anaemia and thrombocytopenia.

Sequence 400 AA;

Query Match 29.8%; Score 711; DB 19; Length 400;

Best Local Similarity 43.7%; Pred. No. 5.9e-56; Mismatches 129; Indels 12; Gaps 5;

108 SDCDIYOTLRGYAOKLSKEKSPFIAYSLVHKDAIMVERLIHAIYNOHNIYCHYDRK 167  
 70 ssckeyltgshytlaptskeadflaylmvlnhfdtfralfaylmpnlycvhndek 129  
 168 APDTFKVAMNNLAKCFENITASKLEAVEYAHISRLQADLNCSDLKSSIQWKYVINC 227  
 130 atefkdvaeqliscfnafiaskepvvygslrqadlnclrdlsafevskvynlnc 189  
 228 GDDPLSNFELVSELKINGANMLFTVKKPNSKLERFTY-NHDLRRVPEYVKLPRTN 286  
 190 ggdiprktnkeivgylgfgsknltpgvlpahalgtrkyvhqehlgkelsv---lrrt 246  
 287 ISKCAPPHNIQIFVGSAYFVLSQAFVKYIENNSIVODFFAKSKDTYSPDEHFWATLIRP 346

Db 247 alkpppphnllyfgsayalerefanfvlndpravdllygskdtfspdenfwylrrip 306  
 QY 347 GIPGELSRSAQVSDLOSTRIYKWNYYEGFFPSCGSHLSVCIYGAELRMLIKDGH 406  
 Db 307 gvpssmpnas-----wtgnllraikwsdmed-rngcnghyngiclygnngdlkwlvnsp 360  
 QY 407 WFANKFDSKVDPIILIKLAELKEEDOR 433  
 Db 361 lfankfelntlypltyecl-elrhrer 385

Search completed: August 2, 2001, 14:44:50  
 Job time: 121 sec

Fri Aug 3 09:38:08 2001

us-09-645-392-2.rag

---

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: August 2, 2001, 14:43:34 ; Search time 25.93 Seconds  
(without alignments)  
1330.778 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389

Sequence: 1 MKIFKCYFHHTQOKYFILF.....DWITLPSEKLFMDRLTTTS 453

Scoring table: BLOSUM62 Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	863.5	36.1	428	A46293	beta-1,3-galactosyl
2	711	29.8	400	A46297	beta-1,6-N-acetylgl
3	326	13.6	445	T21262	hypothetical prote
4	323	13.5	425	T20207	hypothetical prote
5	315.5	13.2	454	T21261	hypothetical prote
6	296	12.4	401	T24929	hypothetical prote
7	294	12.3	472	T2188	hypothetical prote
8	275	11.5	402	T24930	hypothetical prote
9	245	10.3	459	T24742	hypothetical prote
10	229.5	9.6	454	T25382	hypothetical prote
11	220.5	9.2	489	T24013	hypothetical prote
12	219.5	9.2	470	T32137	hypothetical prote
13	219.5	8.5	486	T21566	hypothetical prote
14	203.5	8.5	472	T23384	hypothetical prote
15	202.5	8.5	753	T24745	hypothetical prote
16	197.5	8.3	448	D85042	probable glycosyl
17	193.5	8.1	478	T27714	hypothetical prote
18	188	7.9	467	T21406	hypothetical prote
19	187	7.8	447	T20096	hypothetical prote
20	182.5	7.6	513	T20068	hypothetical prote
21	178	7.5	367	T28892	hypothetical prote
22	175.5	7.3	406	F96571	hypothetical prote
23	168	7.0	395	B96735	unknown protein P2
24	166.5	7.0	434	T51450	probable RING zinc
25	132	5.5	630	T02524	hypothetical prote
26	131	5.5	384	T08940	hypothetical prote
27	126	5.3	371	T21818	hypothetical prote
28	116	4.9	895	E64431	UDP-glucose 6-dehyd
29	108.5	4.5	661	S50734	WH13 protein - yea

30	108	4.5	702	2	E69498	hypothetical prote
31	107	4.5	717	2	S78177	hypothetical prote
32	107	4.5	1072	2	T37742	serine threonine-p
33	107	4.5	2265	2	T26183	hypothetical prote
34	106.5	4.5	1347	2	T41321	BTB domain and Ank
35	104	4.4	364	2	T48628	hypothetical prote
36	104	4.4	2485	1	H71621	serine/threonine-s
37	104	4.4	2628	2	S59413	probable membrane
38	103.5	4.3	588	2	H82877	hypothetical prote
39	102.5	4.3	415	2	I65223	heparin-binding fi
40	102.5	4.3	707	2	A54846	fibroblast growth
41	102.5	4.3	707	2	A38429	keratinocyte growt
42	102.5	4.3	769	2	S16236	fibroblast growth
43	102.5	4.3	820	2	S17295	fibroblast growth
44	102.5	4.3	821	1	TVMSBK	fibroblast growth
45	102.5	4.3	821	1	TVHDF2	fibroblast growth

## ALIGNMENTS

## RESULT 1

A46293  
beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase  
C:Species: Homo sapiens (man)  
C:Date: 20-Oct-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: A46293  
R:Biernutzen, M.F.; Fukuda, M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992  
A:Title: Expression cloning of a cDNA encoding UDP-GlcNAc:gal beta 1-3-galNAc-R (GlcNAc)  
A:Reference number: A46293; PMID:93028457  
A:Accession: A46293  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-428 <BIEM>  
A:Cross-references: GB:97347; NID:q183440; PIDN:AAA35919.1; PID:q183441  
A:Note: sequence extracted from NCBI backbone (NCBI:115900, NCBI:115901)  
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	36.1%	Score	863.5	DB 2:	Length	428:			
Best Local Similarity	43.1%:	Pred. No.	6.2e-56:						
Matches	187:	Conservative	72:	Mismatches	132:	Indels	43:	Gaps	12:
QY	28	LLKLLNRRRL--PQKDIYLV-EYSISTSPVRRKTHVDE--VRY-----EV 71							
Db	1	MLRLLRRRLFSPTKYYFMVLVSLTFVYLR--IHQPEEVSVHHLGAPSSDI 57							
QY	72	NCSGIYEOEPLDIGK-----SLEIRRDITLDEDDVAMTSDCDIYQTLRGYAKLV 125							
Db	58	NCTKVLQGDVNEIQKVKELITVFKKRP--RMTPDYINMTSDCSSFIRKRYIVPLS 115							
QY	126	KEESFPYLVVYHKDAIMVERLIHAIYNOHNIYCHIDPKADITKRVAMNNAKQFSN 185							
Db	116	KEEEFPAYSIVVYHKIEMLDRLIRAIYMPQNVCHVDKSDSYSLAAMGASFSN 175							
QY	186	IFIAKLEAVYAHISRLQADLNCLSDILKSSIQMKYVINCQDPFUKSFTELYSELK 245							
Db	176	VFVASRLSESVYVYASMSRQADLNCKMDLYAMSNMXYLINCGMDPIKTLTEIVRKL 235							
QY	246	LNGANMLETVKRPNSKLERFTYHHELRVPEYV--KIPITNSKRAPPNINIOIFGSA 303							
Db	236	LMGEHNLETERMPHSKHEERW-----KKRAYVNGKL-TNTGTVKMLPLETFLEFGSA 287							
QY	304	YFVLSQAFVYKTIFFNSIYQDFEAFMSKDTYSPDEHFWATLIVPGIPGHSRAO-DVSD 362							
Db	288	YFVSREIVGYLVQNEKTKQKLMENADITYSPDEIVMAIQIIPVPGSLPASHKYDLSDM 347							
QY	363	QSKTRLVKWNVYEGFF-----YPSCTGSHLSRVCICYGAELRWLTKDCHWAFNFKDSKV 417							
Db	348	QVAFRFVQKQVFEEDVNSKGADYPPCDGIVHRSVCIFGAGDLNMMLRKHHLFANKEEDVD 407							

QY 418 PILIKCLAEKLEEQ 431  
 Db 408 LFAICLCDEHLRHK 421

## RESULT 2

A46297  
 beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.-) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
 C:Accession: A46297  
 R:Bierhuizen, M.F.; Mattei, M.G.; Fukuda, M.  
 Genes Dev. 7, 468-478, 1993  
 A:Title: Expression of the developmental I antigen by a cloned human cDNA encoding a mem  
 A:Reference number: A46297, MID:93194065  
 A:Accession: A46297  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-400 <E1>  
 A:Cross-references: GB:219550; NID:q296531; PIDN:CAA79610.1; PID:q296532  
 A:Note: sequence extracted from NCBI backbone (NCBIN:127166, NCBI:127167)  
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 29.8%; Score 711; DB 2; Length 400;  
 Best Local Similarity 43.7%; Pred. No. 9e-45;  
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SDCIYOTLGYAOKLSEKSPFIAYSLVHKDAIMVERLLHAIYNOHNIYCIYHDKR 167  
 Db 70 SSCSEYLTQSHYITAPLSKEADDFLAYIMVHHHPTFARLFRALYMPQNIYCVHDEK 129  
 QY 168 APDTFKYAMNNLAKCFENITASKLEAVEYAHISRLADINCLSLDKSTOKKYIYINLC 227  
 Db 130 ATTEFKADAVQQLSCFPNAPLAKMEPVYVGGISRLQADINCRDLSAFESWRYVINTC 189  
 QY 228 GQDFPLKSNFELVELKLGANLLETKPNSKLERFTY-HHELRYVRYEYKLPRTN 286  
 Db 190 GQDFPLKTNKEIYVGLGFKGNITTPGLPRAHIGTKRVHGHJLKELSYV---IRTY 246  
 QY 287 ISKRAPPHNIOIFVGSAYFVLSQAFVYIFNNSTVODFPMSKDTYSPDEHMAATLRVP 346  
 Db 247 ALKPPPHNLTIVYGSAYVALSRFANFVYHDPRAVDLLQMSKTFSPDHFVYTLNRIP 306  
 Db 307 GYPSMNPAS-----WTGNLRAIKMSDMED-RHGGCHGHVHGCIYGNQDLKWLVSFS 360  
 QY 407 WFANKFDSKVDPIILIKCLAEKLEEQOR 433  
 Db 361 LFANKFELINTYPLTVECL-ELRHRER 385

## RESULT 3

T21262  
 hypothetical protein F22D6.11 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T21262  
 R:Wilkinson, J.  
 submitted to the EMBL Data Library, April 1996  
 A:Reference number: 219397  
 A:Accession: T21262  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-445 <W12>  
 A:Cross-references: EMBL:T21262; PIDN:CAA95817.1; GSPDB:GN00019; CESP:F22D6.11  
 A:Experimental source: clone F22D6  
 C:Genetics:  
 A:Gene: CESP:F22D6.11  
 A:Map position: 1  
 A:introns: 65/3; 95/2; 136/3; 180/3; 256/3; 281/3; 403/3

Query Match 13.6%; Score 326; DB 2; Length 445;  
 Best Local Similarity 25.1%; Pred. No. 2e-16;  
 Matches 108; Conservative 85; Mismatches 176; Indels 62; Gaps 15;

QY 25 LSLKLNVRLRLEPKDIYLVESLSTSPFVRRRYTHVQDEV--RYVNCSGIY----E 78  
 Db 17 LVIFLYLFIVKPLVPR--ILESLNSRNPQETSLSKIEDLLDDINDINLNFNGSKN 73  
 QY 79 QEPLEIGSLERRDIDLEDVAMTSPCDIYQTLRGYAOGLVSEKSPFIAYSLV 138  
 Db 74 RQQLIVNSRST-----EDKLYSTDQCQLTKSLFRNNKPLPSEESFPLSGLL 124  
 QY 139 VHKD--AIWVER--LIHAIYNOHNIYCIYHDKRAPDTFKYAMNNLAKCFENITASKLE 193  
 Db 184 PIDWGSHEITINAYDCLFSLHLSKSDMYRFQYLSGVDLPLTNLEMQOILKHLNG----- 238  
 QY 254 TVKPNNSKLERFTYHHELRYVRYEYKLPRTNISKRAPPHNIOIFVGSAYFVLSQAFVK 313  
 Db 239 -----TANVEIK--PYQYQL--RQKNETQSP--LPLFKSLISLIPREAN 279  
 QY 314 YFPNNSIYODFPAMSKDTYSPDEHMAATLI---RVPQIGPISISADYSDLOSKTRLV- 369  
 Db 280 HLSSSIPQQLLEFLRNGLADEGFWGLFGKNKLFDFGSLINKEWISYNNVETNLTY 339  
 QY 370 ---KNVY---EGFFYPSCTGSHRSYCIYGAELMRLIKDGHFANKFDSKVDPIILIK 422  
 Db 340 PIDGWRVYISRQGISKPNCHMYMKAGSCVRSIGVPLLSKALVAKFKLKSPEAYF 399  
 QY 423 CLAEKLEEQOR 433  
 Db 400 CL---LKEHRR 407

## RESULT 4

T20207  
 hypothetical protein C54C8.11 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T20207  
 R:Dobson, R.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: 219237  
 A:Accession: T20207  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-425 <W12>  
 A:Cross-references: EMBL:T20207; PIDN:CAB05469.1; GSPDB:GN00019; CESP:C54C8.11  
 A:Experimental source: clone C54C8  
 R:McClay, R.  
 submitted to the EMBL Data Library, October 1998  
 A:Reference number: 220372  
 A:Accession: T27478  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-425 <W12>  
 A:Cross-references: EMBL:AL032664; PIDN:CAA21769.1; GSPDB:GN00019; CESP:C54C8.11  
 A:Experimental source: clone Y91F4A  
 C:Genetics:  
 A:Gene: CESP:C54C8.11  
 A:Map position: 1  
 A:introns: 29/2; 117/3; 161/3; 195/3; 237/3; 266/3; 295/1; 313/3; 355/1; 374/2

Query Match 13.5%; Score 323; DB 2; Length 425;  
 Best Local Similarity 22.6%; Pred. No. 3.1e-16;  
 Matches 102; Conservative 85; Mismatches 176; Indels 88; Gaps 14;  
 QY 16 VFILFLTWLLSLKLLNVRLRLE-PQKDIYLVESLSTSPFVRRRYTHVQDEVRYVNC 74

```

Db 5 1FLPFLIFETFDLVLVILEIFESQNPASTIDES-----HECQ 44
QY 75 GYEOEPLIEIGKSLERIRRDIDLEDDVYAM-----TSDCIYQTLRGYAQKLVSK 126
Db 45 KMFQGDVSLA-----RGALFKFDREIILAOILKLSGEOACAEFOKIFGFEQPTSQ 98
QY 127 EKEFPFLAYSIVHKDAIMVERLHAITYNQHNYCHYDKKAPDTEFVAMNNLAKCSNI 166
Db 99 EELFEPFLAYGLVHGDVQSLLSLAIYQPNQFCLAVDNGSNVEFTGLVPMSCYGN 158
QY 187 --FLASKLEAVEYAHISRLQADINCLSDLKSSIQWKYVINCQDPEPLKSNFELVELK 244
Db 159 QYFILDRLMCGYELLTSV---FOCVYLAFLPSDMKYFOYLSGVDAPLKSNLEMIRLK 215
QY 245 KLGANMLIETKPPNSKLERETTYHHELRVPEYEVKLPINISKEAPHNIQIFVCSAY 304
Db 216 ALNG-----SFNAEI--LPFEFYRLNRKRPWSSPLPKTSL--SAT 253
QY 305 FVLQSAFYKTYFNNSIVODFPAMSKDTYSPDEHFMATLIRVPGI-----PGEISRSA 356
Db 254 F--SRKSNFVNSEKYLEQDIFLRTGTCADSEIMATLIGAPKVIKFSLEPMGEPDACA 311
QY 357 QDVSDL-QSKTRLYK-----NYEGFPY-----SCGSHLRVSCITYGAELRW 400
Db 312 WIKNKYRTRGKLGKYGGENQKIDNGYVSRYOQYVNRAPVKCKGYRRLSCVFGYDLPN 371
QY 401 LKDGHWFANKFDSKVPDILIKLAEKLEQ 431
Db 372 LINHELVAKHLIYSYQPAAMCIVENSROK 402

```

## RESULT 5

```

T21261
hypothetical protein F22D6.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21261
R:Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19397
A:Accession: T21261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <WIL>
A:Cross-references: EMBL:Z71262; PIDN:CAA95816.1; GSPDB:GN00019; CESP:F22D6.12
C:Genetics:
A:Gene: CESP:F22D6.12
A:Map position: 1
A:Introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3

```

```

Query Match 13.2%; Score 315.5; DB 2; Length 454;
Best Local Similarity 24.3%; Pred. No. 1.2e-15;
Matches 111; Conservative 79; Mismatches 190; Indels 77; Gaps 16;
QY 8 EKHITLQAKYFLIFLMLSLKLNLNVRRLFPQKDIYLVESISLSPVNRNRYTHKDEV 67
Db 6 FETSLIAIFLFI-----YFVESLIFPRKQ-----EDKNVSKOLKSLCTIASDSY 52
QY 68 ---REVNCSGIEQDEPLEIGKSLERIRRDIDLEDDVYAMTSDCIYQTLRGYAQKLV 124
Db 53 LADNNEINCSNII--KGYKTNEKLDIMHDI--EOLFSCQTNKCOFLKTLFRENTNPM 107
QY 125 SKEEKSFPILAYSLVYHKDA-----IMVERLHAITYNQHNYCHYDKKAPDTEFVAMNNLAKCSNI 166
Db 108 SAEEKFEPFLSYGMLYKDIPOVTPARAMFTKLFLFLINLOVLEFLSSITVHPQNYCIAVGE 167
QY 167 KAPDTEFVAMNNLAKCSNIFLASKLEAVEYAHISRLQADINCLSDLKSSIQWKYVINC 226
Db 168 NSAPIFQNLIREVSTCFSSVNHF-MKRPPISMGSHIIDSVDCLFELSLHLETDWRYFOYL 226

```

```

QY 227 CGDPEPLKSNFELVELKLNKANMLIETKPPNSKLERETTYHHELRVPEYEVKLPIN 286
Db 227 SGVDIPLKTNLEKVOILKLNSTNVEIT---NYQARLTGNE----- 267
QY 287 ISKEAPHNIQIFVGSAYVLSQAFKYIFNNSIVODFPAMSKDTYSPDEHFMATLIRV 346
Db 268 --NESP---LPLEKSSLSAIIIPRKANQLASNTARKLLEFLMNTIEIDEGFWGTLPGNK 322
QY 347 ---GIGET-SRSADQVSLQSK--TRLVKMNY---EGEFYPSCTGSHLR--SVCIYGA 396
Db 323 DQFNISGSINSDKMEYRDNNQNIENPTDGMSTIYSRDQIMPELCKNMYKDCSCVF 382
QY 397 ELRFLIKDGHWFANKFDSKVPDILIKLAEKLEQ 433
Db 383 DVPRLRTSKALVAHKFYLKSEPEAYFCL--LKEHHR 416

```

## RESULT 6

```

T24929
hypothetical protein T15D6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24929
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19956
A:Accession: T24929
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-401 <WIL>
A:Cross-references: EMBL:283125; PIDN:CAB05620.1; GSPDB:GN00019; CESP:T15D6.2
C:Genetics:
A:Gene: CESP:T15D6.2
A:Map position: 1
A:Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3

```

```

Query Match 12.4%; Score 296; DB 2; Length 401;
Best Local Similarity 23.4%; Pred. No. 2.8e-14;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;

```

```

QY 61 THYKDEVRYEYVNCSGIEQDEPLEIGKSLERIRRDIDLEDDVYAMTSDCIYQTLRGYA 120
Db 34 TDLIDL--EINCTNII--QGLNKEQLELINTKMI---EDKLMNSTDRCHTVKSMFRFN 86
QY 121 QKLVSEKESFPPIAYSLVHKDAIMVERLHAITYNQHNYCHYDKKAPDTEFVAMNNLA 180
Db 87 EYPLSEEEARFPISYGLVRELQVLEMLSIYQPNQECIAGVENSASTFILLLELS 146
QY 181 KCFGNIFASKLEAVEYAHISRLQADINCLSDLKSSIQWKYVINCQDPEPLKSNFELY 240
Db 147 DCFPN-----KRPPIMGSEIINSYVDCLEKFLSHLKNMKRYQYLSGVDPILKTMLEMY 201
QY 241 SELKLGANMLIETKPPNSKLERETTYHHELRVPEYEVKLPINISKEAPHNIQIY 300
Db 202 RILKSLNGTANVEIKRYENRRL-----LGQNETESPPLFK 237
QY 301 GSAFYLSQAFKYIFNNSIVODFPAMSKDTYSPDEHFMATLIRVPGI-----IPGEISRSAQ 357
Db 238 SLSLSILPRKAAVLAASSIIPQOLLEFLRMTVWADGEFWGTLPNGNGLEFVPGSLN----- 293
QY 358 DVSDQSKTRLYKWNYYEGFPYPSCTGSHLRVSCITYGAELRWIKDGHWFANKFDSKYD 417
Db 294 -----FEEHQIFESGCHNMKRDGSCVFQIGVSNLLQALVALAKLYLTSE 340
QY 418 PILIKCL 424
Db 341 PEAYFCL 347

```

RESULT 7  
T22188

hypothetical protein F44F4.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T22188

R:Colles: L.

submitted to the EMBL Data Library, September 1994

A:Reference number: Z19528

A:Accession: T22188

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-472 <WIL>

A:Cross-references: EMBL:Z37092; P1DN:CAA85457.1; GSPDB:GN00020; CESP:F44F4.6

A:Experimental source: clone F44F4

A:Gene: CESP:F44F4.6

A:Map position: 2

A:Introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

Query Match 12.38; Score 294; DB 2; Length 472;

Best Local Similarity 21.88; Pred. No. 4.9e-14;

Matches 103; Conservative 83; Mismatches 184; Indels 102; Gaps 17;

32 LNVR-----LPQDIYL-----VEYSLSPFVRNRVTHVDEVRVENCSCGYE 78

24 MNVRFKISQFLYSFTVYVLRGFDYSL-----VPENRK-----LE 63

79 QEPLEIGKSLD--IRNDI-----IDIEDDVVAMSDCDIYQTLGVAQKLVSK 127

64 NIPLECNVINGTENRRKISRAROMDMHFDVVEHEIFNSRVCSTIDKFNFTIRPSSRE 133

128 EKSPPIAVSLVHKDAIYVERLHAIVNQHNYCIYHDKRAPDFFKAMNNLKCFEN-- 185

124 EAEPLAAGLVYTYTYVLTQMSLFYQPMFCITVDOSPNRYKIVIALPSCFFNMH 183

186 IFASKLEAVEYAHISRLQADLNCSDLLKSSIQKRYVINLCGDFPLKSNFELVSELR 245

184 VFEG---EPSQMGFFGLTKNVTCTFNMLSKSKQKRYQYLSGTDLPIRTNLEVRLEFKA 240

246 LKGANMLETVAPPNKLERFTYHHELRVPEYVKLPFRINISKEAPPHNIQIVGSAVF 305

241 LKNGS-----AMTDVSTF-----EVDRY-----KMEGVLEPP--MPYKSSMSV 276

306 VLQSAFYKYIFNNSIVQDFAMSKDYSPDEHFNATLIRVPG---IPGEISRSADVSD 362

277 VVPEGADYLLISPRVQKLKYLKSTWIPDESFMSIVLGSPALLPVGSI--RVRDILML 334

363 QSKTRL-----VKNVYEGF---EYPSCTGSHLSVCITYGAELRWLIKDGHWF 408

335 KKNRKLRRPYENTVNSTISTIGRYQVWGWGKCEFGKVKDFSCVGYEDIEINTRELV 394

409 ANKEDSKVDPILIKCLAEKLEEQ-----RDWITLPESEKLFMDRNL 450

395 AHKILYEQPAFAMCMLEKVRRRSLSPDAHLFSARSYSQMTVELYOGKAIT 446

RESULT 8

T24930

hypothetical protein T15D6.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T24930

R:Robson, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19956

A:Accession: T24930

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-402 <WIL>

A:Cross-references: EMBL:T28125; P1DN:CA805621.1; GSPDB:GN00019; CESP:T15D6.3

A:Experimental source: clone T15D6

A:Gene: CESP:T15D6.3

A:Map position: 2

A:Introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

A:Map position: 1

A:Introns: 55/3; 85/2; 121/3; 161/3; 237/3; 262/3

Query Match 11.58; Score 275; DB 2; Length 402;

Best Local Similarity 24.08; Pred. No. 9.7e-13;

Matches 101; Conservative 69; Mismatches 169; Indels 82; Gaps 14;

16 VFILFTLLSLKLLNLRRLFPQKDIYLYEYSTSPFVNRNRYTHVDEVRVENCSCG 75

19 IFILFIAKALIRPLENPNQRIRHPTDLL-----DDL--QINCTN 56

76 IYQEPLEIGKSLERRRDIDLEDDDVVAMSDCDIYQTLGVAQKLVSKNEKSPFAY 135

57 ILQGFENE--NLEELINRMT--ENKMNWSTDRQTLTSMRFKVPSEARFUSF 111

136 SLVHKDAIYVERLHAIVNQHNYCIYHDKRAPDFFKAMNNLKCFENIFASKLEAV 195

112 GLLYVEKLSQVLEFLSSITQPNQENCIANGENSARAFILILKELANCFN-----KRPPI 166

196 EFAHSRLQADLNCSDLLKSSIQKRYVINLCGDFPLKSNFELVSELRKLGANMLETV 255

167 KNGSEFIIINSVYCLEFSLHLSKMDKRYQYLSGVDIPKTNLEWRILKRLNG-----TV 221

256 KPNKSLERFTYHHELRVPEYVKLPFRINISKEAPPHNIQIVGSAVFYLSQAIFYVI 315

222 NIGIS-----TYEDRL-----LNGKNTESP--LPFKSSLSLIPRAANYL 262

316 FNNSIVQDFAMSKDYSPDEHFNAT-----LINVPDGEISRSADVSDQSKTRLY 369

263 SSSSVYQQLLEFLRTTWVADGEFGWGLFGNKDLFVYVPGSFNNDPLTNGMGVYRHL 321

370 KM-----NYEGGFYPSCTGSHLSVCITYGAELRWLIKDGHWFANKFDSKVDILK 423

322 -WVSECHMYMD-----NS-CYFGIGDPNLMKSAVAHMLYIESEFEARFC 368

424 L 424

369 L 369

RESULT 9

T24742

hypothetical protein T09E11.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T24742

R:McClay, K.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19930

A:Accession: T24742

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-459 <WIL>

A:Cross-references: EMBL:T81147; P1DN:CA803536.1; GSPDB:GN00019; CESP:T09E11.9

A:Experimental source: clone T09E11

A:Gene: CESP:T09E11.9

A:Map position: 1

A:Introns: 39/1; 78/3; 204/1; 291/1; 388/3; 371/3; 417/3

C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 10.38; Score 245; DB 2; Length 459;

Best Local Similarity 22.88; Pred. No. 1.9e-10;

Matches 106; Conservative 74; Mismatches 202; Indels 82; Gaps 17;

43 IYLVEXSLSTSPFVRNRYTHVDEVRVENCSCGY--EQEPLEIGKSLERRRDITLED 100

25 VYIYMFSS-SKSLGRN-----EONSVALYMLNRENTQRFALINELRRDPAKIES 72

101 -----DDVYAMTSDCDIYQTLGVAQKLVSKNEKSPF----- 132

Db 73 KIEHFKYRRRSELAHVDCGRILSGDKDYLTQVSG--ENRIPVEMPNDMSCAVMD 130  
 QY 133 -----IAVSLVHKDAIIVERLHAITYNOHNYCIHDKRAPDFFKAMN 177  
 Db 131 RIITHDHLRLPLKGVAFARIYVDYELVEKQVMSYHPQSPFALDKAPTFKQMR 190  
 QY 178 NLACFSNIFIAKLAEVAHISRLQADLNCISDLKSIQKYYINLGGDFPLKSNF 237  
 Db 191 AMAACLPVLLLPDQEPIDSGHNVNLAHLNCLRALINKP--GNNYAMLLQNHDLTTSVY 249  
 QY 238 ELVSLKLNAGNMEETKPPNSKLEPRTYHEHLRRVYEVKLPITRNISKEAPPHNIO 297  
 Db 250 ELEOVYEWLGGANDVELL--PEAQRLDEENFKMD----PRSLKMPDESKVDETILNKKIK 304  
 QY 298 IFVGSAYVLSQAFVKYIFNNISIVODFF--AMSKDTSPDEHFMAFL--IRVPGIPELISR 354  
 Db 305 FSKGSGVQGSMSRAAVDMKTARKYNLSTYIDQNNQGMVDEMILISLOISAFLGMPGHE-- 362  
 QY 355 SAGQVSDLOSRTLVKMYNVEGFEYPSCTGSHLR--SVCTIGAELRWLIKQGHMFANKFD 413  
 Db 363 TDQCKEGEPPRMVQWMEDESY---CASKYKRNHVCILGLEYLRSAVSFTILMFNKM 419  
 QY 414 SKYDPLIKLCLAEKLEQ---QRDWILPSEKLEPMDBNLTTTS 453  
 Db 420 PFDNSITICTAELLNRTFMGQNDH---PLEEYY--KNMVTVS 459

## RESULT 10

T25382  
 hypothetical protein T27F6.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
 C:Accession: T25382  
 R:DoBson, R.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z20026  
 A:Accession: T25382  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-454 <MIL>  
 A:Cross-references: EMBL:282060; PIDN:CAB04881.1; GSPDB:GN00019; CESP:T27F6.1  
 A:Experimental source: clone T27F6  
 A:Genetics:  
 A:Gene: CESP:T27F6.1  
 A:Map position: 1  
 A:Intons: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 9.6%; Score 229.5; DB 2; Length 454;  
 Best Local Similarity 22.7%; Pred. No. 2.5e-09;  
 Matches 103; Conservative 68; Mismatches 202; Indels 81; Gaps 19;

QY 17 FILFLMLSLKLINVRRLFPQKDIYEVSLSTSPVRRNRYTHVKDEVREYVNCSGI 76  
 Db 10 FIL-----LVALDCLN-----WFLYNGV-----YRNHYWGYTONSSPQEDIAKF 50  
 QY 77 YEDEPLEIGSLERRDDIIDLDDV---VAMTSCDIQYTLRG-YAOKLVKKEKSP 132  
 Db 51 YPTSNKDV---FYRRRP---ETENYNGQVLAGDTAYLKYTGERTIAENESLNS 103  
 QY 133 -----IAVSLVHKDAIIVERLHAITYNOHNYCIHDKRAPDT 171  
 Db 104 CEAMVDKILSRDHLRLPLGVAFARIYVDYELIEKHVMSYHPQSPFALDKRAKE 163  
 QY 172 FKVAMNLLACFSNIF-----ASKLEAVEYAHISRLQADLNCISDLKSIQK 222  
 Db 164 FKERMQAMASCLPVLLPGRFFKNPIHDLSDSHGHNVL--AHYNCLRALINKP--GWN 221  
 QY 223 VINCGODPLKSNFELVSELKLNAGNMEETKPPNSKLEPRTYHEHLRRVYEVKLP 282  
 Db 222 AILLQNHDLITKSYLEKIFNMLGGANDV-AIRPELGRDK-----XHFKMDPMSLK 273

QY 283 IRTNISKAP-----PHNIOIFVGSAYVLSQAFVKYIFNNISIVODFF--AMSKDTSPDEHF 338  
 Db 274 LFNENEIDPVLINTLTKFAKAGVSSLSRAAVDMKTFRVDTLTITIDQNNHETIYVDDQF 333  
 QY 339 WAT--LIRVPGIGEIS--RSADVSDLOSRTLVKMYNVEGFEYPSCTGSHLR--SVCTIYG 394  
 Db 334 TQFQISDFLGMGPHFTDCKIKGITEGITRPAQWTHD---QSKCAKSKSRHGICIMG 390  
 QY 395 AAEIRWLKDGWFANKFDSKVDPIILKLAEL 428  
 Db 391 IEHLSMAKSEHLMFNKVLPLFDYSITICTAELL 424

## RESULT 11

T24013  
 hypothetical protein R07B7.6 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T24013  
 R:Harris, B.  
 submitted to the EMBL Data Library, July 1996  
 A:Reference number: Z19830  
 A:Accession: T24013  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-489 <MIL>  
 A:Cross-references: EMBL:275955; PIDN:CAB00115.1; GSPDB:GN00023; CESP:R07B7.6  
 A:Experimental source: clone R07B7  
 A:Genetics:  
 A:Gene: CESP:R07B7.6  
 A:Map position: 5  
 A:Intons: 42/3; 80/2; 235/3; 287/1; 372/2; 416/1

Query Match 9.2%; Score 220.5; DB 2; Length 489;  
 Best Local Similarity 22.0%; Pred. No. 1.3e-08;  
 Matches 102; Conservative 77; Mismatches 209; Indels 75; Gaps 19;

QY 22 TLWLISLKLINVRRLFP--QKDIYEVSL-----STSPVRRNRYTH----- 62  
 Db 16 TLKLSMSATRYKRLDIDLADINTARVSLQNGKNSPLTRSOEKKPRKNITKMD 75  
 QY 63 ---VDEVRVYVNCSGIYECEPLEIGKSLERRDDI-----LEDDVYAMTSDC 110  
 Db 76 NYFTDDEQF--LNGSEMK-----NNKDVLELYVNNGRMKLDNERLEPLMDC 122  
 QY 111 -DIYOTLRGYAOKLVSKKEKSPPIAYSLVHKDAIIVERLHAITYNOHNYCIHDKAP 169  
 Db 123 PSIKNRIG---DMSFRLKRPPIAFVITYIYELQALISYHDNVNFCFYMSKST 179  
 QY 170 DFFKAMNLLACFSNIFIAKLAEVAHISRLQADLNCISDLKSIQKYYINLGGQ 229  
 Db 180 DLKESVRIKMSCFNIVVYLKREYSLNSGGQDPAHDCDKTLDR--KMDHAILQNF 237  
 QY 230 DPLKSNFELVSELKLNAGNMEETKPPNSKLEPRTYHEHLR--VPEYEVKLPITFNI 287  
 Db 238 DLIRTPOLSDISSELYTSM-----GFDGFSYRYTKAKMPAGKLEFIETGV 290  
 QY 288 SKAPPHNIOIFVGSAYVLSQAFVKYIF---NNSIVODFPAWSKDTSPDEHFMAFL-- 343  
 Db 291 PNEILNRNLIYKSLNEIVYSKVFYKSMFKLMIIILKLD-DNDYGVDEMVLQVTLYE 349  
 QY 344 RVPGIPEL--RSADVSDLOSRTLVKMYNVEGFEYPSCTGSHLR--SVCTIGAELRW 400  
 Db 350 NYLLEGQWESNCTRNHNDI--LRTMTWDSGPGFDEKSKRWKRHGICIMGVEYWE 407  
 QY 401 LIKDGWIFANKFDSKVDPIILKLAEL-----DEQQRDWIT 437  
 Db 408 LKISQGVYANKVMAIFDSTIACMKEMIRNTAGETPNTQWLT 450

## RESULT 12

T32137



hypothetical protein R07C3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T32137

R:Lamar, B.; Kramer, J.

Submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid R07C3.

A:Reference number: 221125

A:Accession: T32137

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1470 <L>AM>

A:Cross-references: EMBL:AF016686; PIDN:AA66233.1; GSPDB:GN00020; CESP:R07C3.3

A:Experimental source: strain Bristol N2; clone R07C3

C:Genetics:

A:Gene: CESP:R07C3.3

A:Map position: 2

A:Introns: 18/1: 78/3; 189/1: 276/1: 319/3: 359/2: 401/3: 441/3

C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 9.2%; Score 219.5; DB 2; Length 470;

Best Local Similarity 21.4%; Pred. No. 1.4e-08;

Matches 95; Conservative 83; Mismatches 207; Indels 59; Gaps 16;

QY 38 FPOKDIYLVEYSLSSTSPFVRNRRTYHKDEYREVNCSGIYEOP-----LEIGKS 87  
 DB 19 FSRNSSYLDPLSLVDRRLAYKRTTEALRTYDSCSEQGFYCKRPETQHVDCGRVLGDK 78  
 QY 88 LEIRRRDIT--LEDDDVVAMTSDCDIYQTLRGYAOQLVSKERKSPFI-----AYSLVH 140  
 DB 79 IQLAFQYVYQSLTSGNRPILENSFLNLCSAISKRIIPKTAQFKRLKNGTAFARIVE 138  
 QY 141 KDAIWEELTAIYNQNIYCIHYDKAPDTFKYAMNNLAKCSNITIAKLAVERAHI 200  
 DB 139 ADYEFIEQVQVASHYPONVFCFALDANSAEFOKMKRLQCCLPNVVLDPVTSYSKGH 198  
 QY 201 SRLQADNLCLSDLLKSIQWYVNLGQDFPLKSNFELSELKLGANMLETVKPPNS 260  
 DB 199 NINLAHYCKMK--LESLRGWGYLMLLQNHVDYTKSYELDRIFELLGAVN-----DV 249  
 QY 261 KLERTHYHELRVPEYVKLPITNISKEPAPNIIQIPGSAFVYLSQAFVKYIFNNST 320  
 DB 250 FMSREIPERRRKHLMKDKLSLKLFRNDKSN--DELEISSVSHSSRAVWKLVEVD 308  
 QY 321 VQDFE-AMSKDYSPDEHFATLIRVPGI--PGEISRSA--QDVSDLSKRLVW--NYE 375  
 DB 309 LSIFFIDQNGTDYGADEGFISTPQMSDLMKPGHFTNDCIHNDIAITYTISLAWPITL 368  
 QY 376 GFYFPCSGSHLRVYCIYGAELRWLTKDGHWFANKEDSKYDPL-----IKCLA 425  
 DB 369 DRAAHCAIGTVRHDCICIGIDFPAI-----SKF---PILNMLPAFVYSIIECTA 415  
 QY 426 EKLEQORDWITLSEKLFMDRL 449  
 DB 416 ELL--HNRTYLGQYDKI--EKNL 435

# RESULT 13

T31566

hypothetical protein F30A10.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T31566

R:Barlow, K.

Submitted to the EMBL Data Library, October 1996

A:Reference number: 219442

A:Accession: T31566

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1486 <WII>

A:Cross-references: EMBL:T81072; PIDN:CAB03022.1; GSPDB:GN00019; CESP:F30A10.4

A:Experimental source: clone F30A10

C:Genetics:

A:Gene: CESP:F30A10.4

A:Map position: 1

A:Introns: 19/3; 59/3; 93/2; 248/3; 300/1; 348/3; 385/2; 429/1

Query Match 9.2%; Score 219.5; DB 2; Length 486;

Best Local Similarity 22.9%; Pred. No. 1.5e-08;

Matches 102; Conservative 80; Mismatches 202; Indels 61; Gaps 17;

QY 15 KYEFLFLMLLSLKLINVR-RLFPKDIY-----LVEXSLSPVRNRRTYHK 64  
 DB 36 KRIFFITALLISSTILTYSKSMIPKFIYRLSKERAPLKHITRSTKRA--DYITE 93  
 QY 65 DEYREVNCSGIYEOPLEIGKSLERRRDITLEDVVAMTSDCDIYQTLRGYAOQLV 124  
 DB 94 SRSRY-LNCARLIDGDVDESIDYVNNGR---MKLDEEKLFQLSMDCDSIQ--NRIFRMP 147  
 QY 125 SKEKSFPIAYSLVYHKAIMWERLIHAIVNHNICHYORKAPDFKYAMNNLACFS 184  
 DB 148 PEKLRPIAFVRNIGYIELOEVFLSYHPDNFYCYAMDSSEKUKSMRTMADCFE 207  
 QY 185 NEFIASKLEAVEYAHISRLQADNLCLSDLLKSSIQWYVNLGQDFPLKSNFELSELK 244  
 DB 208 NTVYLDKREYDMDRAGHKQDAHFCDLKOILDE--HMSHAITLQNFDTLTKPKOLDISE 265  
 QY 245 KINGANML-----ETVAKPPNSKLERFTYHHELRVPE--YVKLPITNISK 289  
 DB 266 ILNTYSIMGEFDYGFYSRRRTFEDMTFAGKTL--FKNE--QSVPLEIILHKLKLSLNE 320  
 QY 290 EAPPHNIQIFGSAFVYLSQAFVKYIFNNSIYQDFEAMSKD--TYSPEHFATLI-RVP 346  
 DB 321 -----VILSKVFVKSLEFKVNLQNVIKRFDRTLTGVDMMMTLENYL 365  
 QY 347 GIPGEISRSQDYSDLSQSTRVLYKNNYBFG--FYFSCGSHR--SVCIYGAELRWLKD 404  
 DB 366 GLDGQWESNC--TVAKEDITLTQTHHLESDGLYQDCKSKMLRHSICVIGVFEOLSKS 424  
 QY 405 GHWFANKFDSKYDPLIKCLAEKLE 429  
 DB 425 PMVIANKVTANFDGFTILCVREMIK 449

# RESULT 14

T33384

hypothetical protein H4IC03.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T33384

R:Attelle, P.; Kramer, J.

Submitted to the EMBL Data Library, July 1998

A:Description: The sequence of C. elegans cosmid H4IC03.

A:Reference number: 221334

A:Accession: T33384

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1472 <LAT>

A:Cross-references: EMBL:AF077545; PIDN:AA26306.1; GSPDB:GN00020; CESP:H4IC03.3

A:Experimental source: strain Bristol N2; clone H4IC03

C:Genetics:

A:Gene: CESP:H4IC03.3

A:Map position: 2

A:Introns: 48/1: 94/3; 117/3; 220/1; 307/1; 354/3; 394/2; 406/1; 442/3

C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 8.5%; Score 203.5; DB 2; Length 472;

Best Local Similarity 20.9%; Pred. No. 2.2e-07;

Matches 90; Conservative 80; Mismatches 189; Indels 71; Gaps 16;

QY 20 FLTLWLSLKLINVRRLF---PQKDIYVEXSLSTSPFVRNRRTYHKDEYREVN--- 72  
 DB 19 FKRIWILLFLAVCFVAYILMFPKGIY---DPLTTDQNGNPTIYQVYVNDLRNELDQKN 75

```

QY 73 -----CSGIYEQP-----LEIGKSLIERRRDIIDLEDDVYAMTSDC 110
DB 76 AIISELQKDFVMEKTERKAIYPRRPETHIDCGRIILADCKVYLQTVSGKDRKIVENK 135
QY 111 DIYOTLRGYAOKLVKEEKSP-----IAYSLVHKDAIMVERLHAIVNHNICIHNR 166
DB 136 NIMMCSAAMNRILPSGNDKRLKNGVAFARIVTYDIEMIEKQVMSHPQNSFCFALDK 195
QY 167 KAPDFFKAMNNLACFSNIFLASKLEAVEVAHISRLQADLNCSDLLKSSIQWKYVNL 226
DB 196 KAPSOFEHLERAMACLPVLLLPDEESYDSAGHNINLAHYNCLFVLINKP-GWNYAAILL 254
QY 227 CGQDPELKSNEFLVSELKLNKANMLEYKPPNSKLEFETYHHELRVRYEVYKPI--- 283
DB 255 QNHDIYKSVYIEIQIYDWLGGANDIE-ITPAGRYDN-----KFKMDPVSLKMERN 306
QY 284 RTNISKAPPHNIQIFVGSAYFVLSQAFVYKIFNNSIVQDFE-AWSKDTYSPDEHFMTL 342
DB 307 ETGIDKNVLTSMKFAKGAOGLSRGAVDMVPTVDLSTYINQNMEGSSGVDEQFIQS- 365
QY 343 IRVP---GIPGEIS-----RSADVSDLOSKTRLVKKNY---YEGFFYP--SQTGSHL 387
DB 366 FQVSADLGRPGHPTDECLKOKRNTDFVS-----RMSQMYGSAHKSFDYSIVECTAELL 419
QY 388 RSVCIYGAEL 397
DB 420 YNRTFLGQVD 429

```

```

RESULT 15
T24745
hypothetical protein T09E11.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24745
R:McLay, K.
submitted to the EMBL Data Library, October 1996
A:Reference number: 219930
A:Accession: T24745
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1753 <util>
A:Cross-references: EMBL:Z81147; PIDDN:CAB03539.1; GSPDB:GND0019; CESP:T09E11.6
A:Experimental source: clone T09E11
A:Gene: CESP:T09E11.6
A:Map position: 1
A:Introns: 34/2; 54/2; 191/3; 275/3; 315/3; 359/3; 380/3; 403/3; 426/3; 530/1; 617/1; 66

```

```

Query Match 8.5%; Score 202.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 4.9e-07;
Matches 96; Conservative 76; Mismatches 207; Indels 61; Gaps 16;

```

```

QY 21 LITLWL-----SLKLINRYRLFPKDIY--LYEYSL-----TSPFVRNRYTHVKDEV 67
DB 335 LNNRLLITIAASTISFYVILLLEKEEPYQTLSESTLSCGEGPTKODLHRHTDIADYV 394
QY 68 R--YEVNCSGIYEQP-----LEIGKSLIERRRDIIDLEDDVYAMTSDCIYQTLRGYA 120
DB 395 RERGFADDRKGSYRRRPFANVDCGRILAGDKPYLQSLTGTRNRYKIVENCNLMNSCKAIR 454
QY 121 QKIVSKEEK-----SPPIYASIVVHKDAIMVERLHAIVNHNICIHNRKAPDFFKVA 175
DB 455 SRLPNSNDNLRPLKHGIAFARIVTYDIEMIEKQVMSHPQNAFCFVIDINASEFFKR 514
QY 176 MNNLACFSNIFLASKLEAVEVAHISRLQADLNCSDLLKSSIQWKYVNLGGQDFPLKS 235
DB 515 MRALACMPRVIVLADBEDPYVSSGHNVNLYHNKCLKALLDIP-GWNYAALLQNHDLIMKS 573
QY 236 NFEVLVSELKLNKANMLEYKPPNSKLEFETYHHELRVRYEVYKLP-PI-----RTNI 287

```

```

DB 574 VYEMEQIFEWLGGANDIEFV-----HEIGRVYKRLKMDPMSKLFINETEM 620
QY 288 SKRAPPHNIQIFVGSAYFVLSQAFVYKIFNNSIVQDFE-AWSKDTYSPDEHFMTLIRVP 346
DB 621 DKLLITPMKIVKGWVCSLSRASVEMMFQKLPDSIFMHQINQGRYGVDEQFPIQANA 680
QY 347 --GIPGEIS--RSADVSDLOSKTRLVKKNYEGFFYP--SC-TGSHLRVCIYGAELRW 400
DB 681 EFGMRGHTDECLQOGKTEFETIRIALM-----VPESKCDINMTRHANCIIGLEHFRQ 733
QY 401 LKDGHWFANKFDSKVDPI 420
DB 734 VASFTHLMFNFKVSSSLDELI 753

```

```

Search completed: August 2, 2001, 14:45:50
Job time: 136 sec

```

Fri Aug 3 09:38:10 2001

us-09-645-192-2.rpt

---

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:45:20 ; Search time 16.71 Seconds  
(without alignments)  
928.650 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389  
Sequence: 1 MKIFKCYRKHTLQKVFILF.....DWITLPSEKLFMDRLTTTS 453

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	863.5	36.1	428 1	G6NT_HUMAN
2	848.5	35.5	428 1	G6NT_MOUSE
3	840	35.2	427 1	G6NT_BOVIN
4	730.5	30.6	400 1	BCIB_MOUSE
5	711	29.8	400 1	BCIB_HUMAN
6	716	4.9	895	YAS4_YEAST
7	108.5	4.5	661 1	WHI3_YEAST
8	102.5	4.3	654 1	BFR2_HUMAN
9	102.5	4.3	821 1	BFR2_HUMAN
10	102.5	4.3	821 1	FGR2_MOUSE
11	101.5	4.2	2054 1	YCF2_PINTH
12	101.5	4.2	2280 1	YCF2_TOBAC
13	100.5	4.2	2607 1	YCF2_TOBAC
14	99.5	4.2	666 1	BACB_BACLI
15	99.5	4.2	823 1	VOO1_VACCC
16	99.5	4.1	823 1	CEK3_CHICK
17	97.5	4.1	1314 1	A23D_DROME
18	97.5	4.1	1314 1	TEFX_CLOTE
19	97.5	4.1	440 1	GAP_HUMAN
20	95.5	4.0	1036 1	YOB6_CAEEL
21	95.5	4.0	936 1	YOB6_CAEEL
22	94	3.9	701 1	MSH4_HUMAN
23	94	3.9	364 1	SYGB_HELPJ
24	94	3.9	454 1	DP3B_MYCCE
25	94	3.9	1344 1	DNAI_BUCAP
26	94	3.9	1675 1	IF3A_MOUSE
27	94	3.9	1675 1	IF3A_MOUSE
28	94	3.9	1675 1	IF3A_MOUSE
29	93.5	3.9	884 1	TEI1_YEAST
30	93.5	3.9	2787 1	TEI1_YEAST
31	93	3.9	503 1	Y208_SHEEP
32	93	3.9	520 1	Y208_METJA
33	93	3.9	576 1	POLO_DROME

34	92	3.9	908 1	DP01_BORNU	OS1498 borrelia bu
35	92	3.9	1103 1	CGF_BOVIN	OS2740 bos taurus
36	92	3.9	1382 1	IF3A_HUMAN	Q14152 homo sapien
37	92	3.9	2022 1	ANT1_ONCHO	P21249 onchocera
38	91.5	3.8	720 1	SPOT_MYCCE	P47520 mycoplasma
39	91.5	3.8	757 1	RTS1_YEAST	P38903 saccharomyc
40	91.5	3.8	1956 1	ATX1_PLAFA	O04956 plasmodium
41	91	3.8	344 1	Y613_METJA	O58030 methanococc
42	91	3.8	458 1	Y613_EUGER	P05728 euglena gra
43	91	3.8	759 1	TEI1_YEAST	P47988 saccharomyc
44	91	3.8	809 1	LEP_BACAN	P15917 bacillus an
45	91	3.8	993 1	NISB_LACIA	P20103 lactococcus

## ALIGNMENTS

RESULT ID	66NT_HUMAN	STANDARD:	PRT:	428 AA.
AC	002742:			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	BETA-1,3-GALACTOSYL-O-GLYCOPROTEIN BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)			
DE	(CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT)			
GN	G6NT.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
ON	NCBI_TaxID=9606;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93028457; PubMed=1329093;			
RA	Bierhuizen M.F.A.; Fukuda M.;			
RT	"Expression cloning of a cDNA encoding UDP-GlcNAc:gal beta 1-3-galNAc-R (GlcNAc to GalNAc) beta 1-6-glcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen."			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96078409; PubMed=7579796;			
RA	Bierhuizen M.F.A.; Maemura K.; Kudo S.; Fukuda M.;			
RT	"Genomic organization of core 2 and I branching beta-1,6-N-acetylglucosaminyltransferases. Implication for evolution of the beta-1,6-N-acetylglucosaminyltransferase gene family."			
RL	Glycobiology 5:417-425(1995).			
CC	- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.			
CC	- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL-1,3-(N-ACETYL-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-GALACTOSAMINYL-R.			
CC	- SUBSTRATE: GLYCOSYLATION.			
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.			
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES AND MYELOID CELLS.			
CC	- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL: M97347; AAA35919.1;			
DR	EMBL: L41415; AAA96651.1;			
DR	PIR: A46293; A46293.			
DR	MIM: 600391;			
KW	Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;			

KM Goldi stack: Glycoprotein.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT  
 FT DOMAIN 33 428 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 428 AA; 49784 MW; 7C9CEB51DA00A60 CRC64;

Query Match 36.1%; Score 863.5; DB 1; Length 428;  
 Best Local Similarity 43.1%; Pred. No. 2e-55;

Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

QY 28 LKLLNVRRLF--PQKDIYLV-EYSLSSTPFRNRYTHVDE--VRY-----EV 71  
 DB 1 MLRLRLRRLFSYPTKYFENVLVLITFEYLK---IHQKEPFSVRHLELAGENSSDI 57  
 QY 72 NCSGIYDEPLEIGK-----SLEIRRDIILEDVVAMTSDCIIYQTLRGYAKLYS 125  
 DB 58 NCTLVQGDVNEIÖKVKLEILTVKFKRP--RMTPDYIMTSDCSSFKRRKYIPELS 115  
 QY 126 KEKSPFIAYSLVYHKDAIWERLIHAIYNOHNYCIHYRRKAPDFEYAMNLLACFSN 185  
 DB 116 KEKSPFIAYSLVYHKDAIWERLIHAIYNOHNYCIHYRRKAPDFEYAMNLLACFSN 175  
 QY 186 IFIASKLEAVEYAHISRLQADLNCSDLLKSSIQWKYVNLGQDPPLKSNFEVSELK 245  
 DB 176 VFVASRLESVYVYASWSVQADLNCMDLAMSANWKYLMICMDPEPKINLEIVKRLK 235  
 QY 246 LINGANMLETVKPPNSKLEFTHYHHLRVRVEYV--KLPIRTNISKAPPHNIQIFVSGA 303  
 DB 236 LMGNNLETEREMSHKDEKRM-----KKREYVNGKL-TNTGTVMKLPLEPLEISGSA 287  
 QY 304 YFVLSQAFVKYIFNNSIYODFFAMSKDTPSPDEHFMTLLIRVQIPGELISRSQ-DVSD 362  
 DB 288 YFVVSREYGVVLQNEKIÖKLMWAQDTYSPDEYLMATIORIPEVPSPLASHKYDLSDM 347  
 QY 363 OSKRILYKMYNVEGEF-----YPSCTGSHLRVSVCIYGAELRWLIKDGHWFAKFSKVD 417  
 DB 348 QAVAFKWKQYFEGDVSQKADYPPCDGVHVSVCIFSGADLMMRLHANKFEDVD 407  
 QY 418 PILIKCLAELKEQ 431  
 DB 408 LFAIQCLDEHLRRK 421  
 RESULT 2  
 G6NT\_MOUSE STANDARD; PRT; 428 AA.  
 ID G6NT\_MOUSE  
 AC Q09324;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-  
 DE ACETYLGALUCOSAMINYTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)  
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT).  
 GN G6NT1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DA/2;  
 RA Warren C.E., Smookler D.S., Dennis J.W.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-  
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-  
 CC GALACTOSYL-1,3-N-ACETYL-BETA-D-GLUCOSAMINYL-1,6-N-ACETYL-D-  
 CC GALACTOSAMINYL-R.

CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).

CC EMBL: U19265; AAA60948.1; -  
 DR MGD; MGI:95676; G6NT1.  
 KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KM Goldi stack: Glycoprotein.

FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT  
 FT DOMAIN 33 428 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 58 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 428 AA; 49649 MW; F3FED2517B27AAB1 CRC64;

Query Match 35.5%; Score 848.5; DB 1; Length 428;  
 Best Local Similarity 42.4%; Pred. No. 2.4e-54;

Matches 183; Conservative 72; Mismatches 138; Indels 39; Gaps 11;

QY 28 LKLLNVRRLF--PQKDIYLV-EYSLSSTPFRNRYTHVDE--VRY-----EV 71  
 DB 1 MLRLRLRRLFSYPTKYFENVLVLITFEYLK---IHQKEPFSVRHLELAGDPPYSNV 57  
 QY 72 NCSGIYDEPLEIGK-----SLEIRRDIILEDVVAMTSDCIIYQTLRGYAKLYS 125  
 DB 58 NCTLVQGDVNEIÖKVKLEILTVQFKRP--RMTPDYIMTSDCSSFKRRKYIPELT 115  
 QY 126 KEKSPFIAYSLVYHKDAIWERLIHAIYNOHNYCIHYRRKAPDFEYAMNLLACFSN 185  
 DB 116 KEKSPFIAYSLVYHKDAIWERLIHAIYNOHNYCIHYRRKAPDFEYAMNLLACFSN 175  
 QY 186 IFIASKLEAVEYAHISRLQADLNCSDLLKSSIQWKYVNLGQDPPLKSNFEVSELK 245  
 DB 176 VFVASRLESVYVYASWSVQADLNCMDLAMSANWKYLMICMDPEPKINLEIVKRLK 235  
 QY 246 LINGANMLETVKPPNSKLEFTHYHHLRVRVEYV--KLPIRTNISKAPPHNIQIFVSGA 303  
 DB 236 STGNNLETEREMSHKDEKRM-----KRYAVVDGKL-TNTGTVMKLPLEPLEISGSA 289  
 QY 306 YFVLSQAFVKYIFNNSIYODFFAMSKDTPSPDEHFMTLLIRVQIPGELISRSQ-DVSD 364  
 DB 290 YFVVSREYGVVLQNEKIÖKLMWAQDTYSPDEYLMATIORIPEVPSPLASHKYDLSDM 349  
 QY 365 KTLVKNMYNVEGEF-----YPSCTGSHLRVSVCIYGAELRWLIKDGHWFAKFSKVD 419  
 DB 350 IAFVAKWKQYFEGDVSQKADYPPCDGVHVSVCIFSGADLMMRLHANKFEDVD 409  
 QY 420 PILIKCLAELKEQ 431  
 DB 410 AIOCLDEHLRRK 421  
 RESULT 3  
 G6NT\_BOVIN STANDARD; PRT; 427 AA.  
 ID G6NT\_BOVIN  
 AC Q92180;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-  
 DE ACETYLGALUCOSAMINYTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)  
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).  
 GN G6NT1.

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Li C.M., Cheng P.W.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-  
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-  
 CC GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-  
 CC GALACTOSAMINYL-R.  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U41320; AAA83244.1; -  
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KM Golgi stack; Glycoprotein.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 32  
 FT (POTENTIAL).  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 52 427  
 FT 58 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 95 95  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 427 AA; 49733 MW; 41B9CBF94BD4196 CRC64;

Query Match 35.2%; Score 840; DB 1; Length 427;  
 Best Local Similarity 42.8%; Pred. No. 9.8e-54;  
 Matches 185; Conservative 73; Mismatches 134; Indels 40; Gaps 12;

QY 28 LKLNLRRL-FPOKDIYL-VEYSLSTSPYRNRYTHVDE-VRY-----EV 71  
 DB 1 MLRKLMRRKLFSEFPTRYFLFLAFGVVFTVLR--IHQKTEFVNGHLEFEENPSSNI 57  
 QY 72 NCSGIYEOBLEIGK-----SLERRRDIDLEDDVAVMTSCDIYOTLRGYAOKLSVE 127  
 DB 58 NCTKILQSDVDLQYKLESLIVKFKRKRAKMTNYINKTGCSAFIKRKYTPPLSK 117  
 QY 128 EKSPFLASLVVHKDAIWERLIIAIVNOHNYCIHYDRKADPTEKVAAMNIAKCSNIF 187  
 DB 118 EAGFLPASYIVVHNHRIEMDLRLAIYMPONFYCIHVAKSEKSLAAVGIASCFSNVF 177  
 QY 188 IASKLEAVYAMISRLQADNLNCLSDLKSSIQMKVIVNCGDPELKNFELYSKELTN 247  
 DB 178 VASQLESVVYASMSRQADLNCMDIYQNMAGKYLINCGMDPIKTNLEIVRKLKLM 237  
 QY 248 GANMLETVKPSKLERFVYHHELRVPEYKLPRTNI--SKDAPPHNIQIFGSAVF 305  
 DB 238 GENNLETEMPSHKKERWKH-----YEVVNGKL-TNMGTDKIHPLPLETFLFGSAHF 289  
 QY 306 VLSQAFVKITFNKSIYQDFANSKDTYSPDEFKATLIRVPIBPEISASQ-DVSDQS 364  
 DB 290 VVSREVEVVLQNOIKQEMWAKDTYSPDEYLAFTIQIPIPEVPSLSIYVYDSDQA 349  
 QY 365 KIRLVVNNYVEGF-----YPSCTGSLRSVCITYGAELRWLIKDGHWANKFDEKVPPI 419  
 DB 350 IARFVWQMFQEGDVSKGAYPPQS-VHRSVCVFAGGLDNLMLHHLFANKFDIDILF 408  
 QY 420 LKCLAELKEBO 431

DB 409 AIOCLDEHLRHK 420  
 RESULT 4  
 BG1E MOUSE  
 ID BG1E MOUSE STANDARD: PRT: 400 AA.  
 AC P97402;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE N-ACETYLGLUCOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE  
 DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)  
 DE (IGNT) (LARGE I ANTIGEN-FORMING BETA-1,6-N-  
 DE ACETYLGLUCOSAMINYLTRANSFERASE).  
 GN GCNT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=9728061; PubMed=9134435;  
 RX Magnet A.D., Fukuda M.;  
 RT "Expression of the large I antigen forming beta-1,6-N-  
 RT acetylglucosaminyltransferase in various tissues of adult mice";  
 RL Glycobiology 7:285-295(1997).  
 CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED  
 CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I  
 CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED  
 CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL  
 CC -1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-  
 CC GLUCOSAMINYL-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U68182; AAB39621.1; -  
 DR MGD: MGI:1100870; Gcnt2.  
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KM Golgi stack; Glycoprotein.  
 FT DOMAIN 1 6  
 FT TRANSMEM 7 25  
 FT (POTENTIAL).  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 26 400  
 FT 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 255 255  
 FT 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 314 314  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 388 388  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 400 AA; 45507 MW; 193FF3862E3B9691 CRC64;

Query Match 30.6%; Score 730.5; DB 1; Length 400;  
 Best Local Similarity 40.0%; Pred. No. 7.6e-46;  
 Matches 156; Conservative 51; Mismatches 160; Indels 23; Gaps 6;

QY 43 ILYVEISLSTSPYRNRYTHVKNDEYEVNCSGIYEDPELIGKSLERRDDIDLEDD 102  
 DB 18 VCVAIVLSEFGDGSYKLNISDSVRSQVCSFID-----GKSRLWKNKL----- 64  
 QY 103 VVAMTSPDDIDYQTLRGAKOLVSKSEKSPFLASYLVVHKDAIWERLIIAIVNOHNYCI 162  
 DB 65 MIHKRPSCTEYVLTQSHITAPLSQEDGDFLAYVYMWVTHHNFDFARLIRAIYMPONFY 124

```

QY 163 HYDRAPDFKVMNNLAKCFNSIFLASKLEAVEYAHISRLQDLNCLSLKSSIQWKY 222
DB 125 HVEKRTAEKGAEDVLSQFPPVFLASKMEPVYGGISRLQDLNCLSLSTSEPMKY 184
QY 223 VINLQSGDEPLKSNFELVSELKINGANMLETVPKPSKLERFTY-HNELRVPEYVKL 281
DB 185 AINTCODEPLKTNKEIVQKLGKKNLTPGVLPRAHAIGRTYVHREHLSKELSY-- 242
QY 282 PITINISKEAPPHNIQIFVGSATFVLSQAFVKYIFNNISIVODEFANSKDTYSPDEHFMAT 341
DB 243 -ITRTALKPPPHNLTITFCSAVALSREFANVLDPRAVDLHWSKDTSPDEHFMAT 301
QY 342 LIRVPGLPGEISSADVDSDQSKTRLVKNNYEGFFYPSCTGSHLSVCIYGAELRWL 401
DB 302 LNIHPGVGSPPMAS-----WTGNLRAVKMDMEK-HKGCGGHYVHGICITGNGDQLWL 355
QY 402 IKDGHFANKFDSKVDPIILKICAEKLEEQ 431
DB 356 INSQSLEFANKFELNTYPLTVECELELRER 385

RESULT 5
BGIB_HUMAN STANDARD; PRT; 400 AA.
ID BGIB_HUMAN
AC 006450;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE N-ACETYLGLUCOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)
DE (ICGT).
GN GCNT2 OR IT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194065; PubMed=8449405;
RA Bierhuizen M.F.A., Mattel M.-G., Fukuda M.;
RT "Expression of the developmental I antigen by a cloned human CDNA
RT encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene
RT family."
RL Genes Dev. 7:468-478(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96078409; PubMed=7579796;
RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT "Genomic organization of core 2 and I branching beta-1,6-N-
RT acetylglucosaminyltransferases. Implication for evolution of the
RT beta-1,6-N-acetylglucosaminyltransferase gene family."
RL Glycobiology 5:417-425(1995).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=98070745; PubMed=9405606;
RA Sasaki K., Kurata-Miura K., Ujita M., Angata K., Nakagawa S.,
RA Sekine S., Nishi T., Fukuda M.;
RT "Expression cloning of cDNA encoding a human beta-1,3-N-
RT acetylglucosaminyltransferase that is essential for poly-N-
RT acetylglucosamine synthesis."
RL Proc. Natl. Acad. Sci. U.S.A. 94:14294-14299(1997).
CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED
CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC -1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
CC GLUCOSAMINYL-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, HIGHLY EXPRESSED IN PROSTATE AND
CC TO A LESSER EXTENT IN SMALL INTESTINE AND COLON. BARELY DETECTED

```

```

CC CC IN HEART, BRAIN, KIDNEY AND PANCREAS. NO EXPRESSION IN PLACENTA,
CC CC LUNG, LIVER, SKELETAL MUSCLE, SPLEEN, THYMUS, TESTIS, OVARY AND
CC CC PERIPHERAL BLOOD LEUKOCYTES. IN FETUS, HIGHLY EXPRESSED IN BRAIN
CC CC AND TO A LESSER EXTENT IN LUNG AND KIDNEY. BARELY DETECTED IN
CC CC LIVER.
CC CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES DRAMATICALLY DURING
CC CC DEVELOPMENT AND ONCOGENESIS.
CC CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (CGNT).
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb.ch/announce/
CC CC or send an email to license@isb.ch).
CC CC
DR EMBL; L19659; AAA81777.1; -
DR EMBL; Z19550; CA97610.1; -
DR EMBL; L41607; AAA99832.1; -
DR EMBL; L41605; AAA99832.1; JOINED.
DR EMBL; L41606; AAA99832.1; JOINED.
DR PIR; A46297; A46297.
DR MIM; 600429; -
DR MIM; 110800; -
DR Transferrase; Glycosyltransferase; Transmembrane; Signal anchor;
KW Golgi stack; 1
FT DOMAIN 1
FT TRANSMEM 7 25
FT FT CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT FT LUMENAL, CATALYTIC (POTENTIAL).
FT FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 26 400
FT FT CARBOHYD 37 400
FT FT CARBOHYD 255 255
FT FT CARBOHYD 314 314
FT FT CARBOHYD 388 388
FT FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 400 AA; 45854 MW; 1469628690A1F43B CRC64;

Query Match 29.8%; Score 711; DB 1; Length 400;
Best Local Similarity 43.7%; Pred. No. 2e+44;
Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SCDIYQTLRGYAKOLVSKESKSPFIAYSLVHKDAIWERITIAIYNQHNITCIHDK 167
DB 70 SSKCEYLQSHITAPLSSEADPFLAYIMVTHHFDFFARLRAIYPMQNIYCVHDEK 129
QY 168 APDTKVMANNLAKCPNSIFLASKLEAVEYAHISRLQDLNCLSLKSSIQWKYVILNC 227
DB 130 ATTERDAVEQLSCFPNFAFLASKMEPVYGGISRLQDLNCLSLSTSEPMKYVNTC 189
QY 228 GDFPLKSNFELVSELKINGANMLETVPKPSKLERFTY-HNELRVPEYVKLPRTN 286
DB 190 GDFPLKTNKEIVQKLGKKNLTPGVLPRAHAIGRTYVHREHLSKELSY--IRTY 246
QY 287 ISKEAPPHNIQIFVGSATFVLSQAFVKYIFNNISIVODEFANSKDTYSPDEHFMATLIRP 346
DB 247 AKRPPPHNLTITFCSAVALSREFANVLDPRAVDLHWSKDTSPDEHFMATLIRP 306
QY 347 GIPGTSRQADVDSDQSKTRLVKNNYEGFFYPSCTGSHLSVCIYGAELRWLNGH 406
DB 307 GYFGSPPMAS-----WTGNLRAIKWMDMEK-HKGCGGHYVHGICITGNGDKMLVNSPS 360
QY 407 WFANKFDSKVDPIILKICAEKLEEQ 433
DB 361 LFANKFELNTYPLTVECL--ELRHRR 385

RESULT 6
YAS4_METJA STANDARD; PRT; 895 AA.
ID YAS4_METJA
AC Q58454;
DT 15-JUL-1998 (Rel. 36, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)

```

```

Query Match          4.9%; Score 116; DB 1; Length 895;
Best Local Similarity 20.1%; Pred. No. 0.67;
Matches 91; Conservative 73; Mismatches 153; Indels 136; Gaps 23.

QY 42 DILVEYLSSTPPFARNRYTHVDEVRVYVNGSGLY--EQEPLGKSLERRRDIDLE 99
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 375 ELIDLE-ELSKDLEIKVHINKDLATNFFNLIKPLSKKYHVDVKNRQITAKDLPLK 433
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 100 DDDVYAMTSDCDIYOTLRGVAOKLYSKEEKSEFFIAYSLVYHND-----AIWERTLHAI 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 434 E-----ILDKY-----GSKNRNLFYAKSKSTPIPKIKIDKFARLIGYISLSEGWISKD 481
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 154 YVQHN-----YCIHRDKRAPDFFKAMNNLACF-----SNFFISKLEAVRY 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 482 YGRNGVRRKRRIGLCEFIHEEYINDV-KNILKLGIKYIEKIKDGHSLISLICKILAYVE 540
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 198 AHISRLQADLNCLSDDL-----KSIOMKYVINLQGDPLKSNFELVSELEKTLNGAN 250
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 541 ENI--LNCINCINCKNRIPOMRNAKEITWEE-----LKGIFRDDGGIVRLNNNAK 588
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 251 MEIYKPPNSKLEFRTYHHELRVRVYEVKPLIRTNISKAPRNHIOIFVGSAYFVLSQA 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 589 -----NLTNEFAT-----VSKKM-AHSLLILQLLGIYAS-- 617
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 311 FVKYIFNNISIVQDFEWMKSDIYSPDEHFMALIRVGP-----GEISRADVD 361
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 618 -YKCCYNN-----KST-----IMAYIRINGLFGQVKKRIGELLFCKKMENTYDIAE 660
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 362 LQSKT-----RLVKWVYEGEFGFPSCGSHLRVLCYGAELFWLIR 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

D6	661	SYKNIEDELGKCKSDNFAILEVEKEIK-EHYSGYVSVSEIENSU-LITSYGILIHNOCFK	718
QY	404	DGHWFANKED-SKYDPTILIKCLAEKLEEQODM	435
		: : : : :       : : : : :	
Db	719	DVKALIKOFENNIEPIITKA-TDIVNEEQIKW	750
<hr/>			
RESULT	7		
ID	WHI3_YEAST	STANDARD;	PRT; 661 AA.
AC	P34761:		
DT	01-FEB-1994 (Rel. 28, Created)		
DT	01-FEB-1994 (Rel. 28, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	WHI3 PROTEIN.		
GN	WHI3 OR YNL197C OR N1382.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Nash R.S., Fletcher A.B.;		
RL	Submitted (Aug-1993) to the EMBL/GenBank/DDBJ databases.		
RC	[2]		
RF	SEQUENCE FROM N.A.		
RG	STRAIN=S288C / FY1679;		
RX	MEDLINE=95242839; Pubmed=7725799;		
RT	Jonniaux J.-L., Coster F., Purnelle B., Goffeau A.;		
RT	"A 21.7 kb DNA segment on the left arm of yeast chromosome XIV		
RT	carries WHI3, GC62, SPX18, SPX19, an homologue to the heat shock gene		
RT	SBE1 and 8 new open reading frames of unknown function."		
RU	Yeast.10:1633-1645(1994).		
SC	- FUNCTION: INVOLVED IN SIZE CONTROL; POSSIBLE RNA BINDING PROTEIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation at		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as their content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb.ch/announce/		
CC	or send an email to license@isb.slb.ch).		
CC	-----		
DR	EMBL; U01095; AAA03320.1; -		
DR	EMBL; X78898; CAAS5511.1; -		
DR	EMBL; Z71473; CAA96092.1; -		
DR	HSSP; F09012; 201A.		
DR	SGD; S0005141; WHI3.		
DR	InterPro; IPR000504; -		
DR	Pfam; PF00076; trm; 1.		
DR	PROSITE; PS50102; RRW; 1.		
KW	DNA-binding.		
FT	DOMAIN 247		
FT	SEQUENCE 661 aa; 71253 MW; 4746B9716CF96043 CRC64;		
	GIN-RICH		
Query Match	4.5%; Score 108.5; DB 1; Length 661;		
Best Local Similarity	20.0%; Pred. No. 1.6;		
Matches	68; Conservative 60; Mismatches 125; Indels 87; Gaps 14.		
QY	117	RGYACKLVKREKSPFIANSLYVHKDAI-WVERLIHALTYNOHNITCIHYDKRAPPTFY	174
		: : : : : : : : : : : : : : :	
Db	52	RGSAGSELVLDNTNSLAISMINTALMSMODIASSNISNH-----DNNIKSYSL	104
QY	175	AMNNLAK-----CFENIFIASRLEAVEY-----AIRSRQAADINCLSD-----LKS	216
		: : : : : : : : : : : : : : :	
Db	105	KLSNAKOTTLKECYAFILAEGVSKIELOKNSSSIISASLEDNDIFIARPELLWL	164
QY	217	STCWKVIVN---LCGDFFPKSNFELVELKK-----LNGANNLETYK	256
		: : : : : : : : : : : : : : :	
Db	165	AIIVAYAILSKNELFPSPPNKTTFVEIIDDTKNLVSFPSSAIFNDTSIRKSN--SGAK	222
QY	257	PPN--SKLRPTY-----HHDLRRPYEVYVLPIRTINISKAPPHNIQIFVGSAV	304



```

Db 223 RPLSLSRBRFRSFPSPNSDPLSQOQQOQPPQPOHSTOKHSPQOCCQOQVNSI 282
QY 305 FVLSQAFVKYIFENNSIVQDFPWS-----KITYSPDEHFATLIRVPGIGE 351
Db 283 PLSSQGVIGLHNSHSHQDLVSVESTIOTSDIGKSPLLRNTEINERKWT-----SIPSS 338
QY 352 ISRSQDVSDLOSKTRLVK-----NYEGFFPSCOTS 385
Db 339 INGV---MSTPQSTPLEMGNTISASQHGSSFLPSAST 375

RESULT 8
BFR2_HUMAN
ID BFR2_HUMAN STANDARD; PRT; 654 AA.
AC 001742;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR BFR-2 PRECURSOR (EC 2.7.1.112).
GN BFR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_Taxid=9606;
RX MEDLINE-91274356; PubMed-1647213;
RA Seno M., Sasada R., Watanabe T., Ishimaru K., Igarashi K.;
RA "Two CDNA's encoding novel human FGF receptor.";
RL Biochim. Biophys. Acta 1089:244-246(1991)
CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X56191; NOT_ANNOTATED_CDS.
DR HSSP: P11362; IFG1.
DR InterPro: IPR000719;
DR InterPro: IPR001245;
DR InterPro: IPR003006;
DR Pfam: PF00047; Ig_2;
DR Pfam: PF00069; Kinase_1.
DR PRINTS: PR00109; TYRKINASE.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW Transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1 21
FT CHAIN 22 654
FT DOMAIN 22 263
FT TRANSMEM 264 284
FT DOMAIN 285 654
FT DOMAIN 57 123
FT DOMAIN 156 234
FT DOMAIN 367 647
FT NP_BIND 373 381
FT BINDING 403 403
FT ACT_SITE 512 512
FT MOD_RES 543 543
FT CARBOHYD 113 113

```

```

FT CARBOHYD 126 126
FT CARBOHYD 150 150
FT CARBOHYD 182 182
FT CARBOHYD 203 203
FT CARBOHYD 214 214
FT CARBOHYD 235 235
SQ SEQUENCE 654 AA; 73594 MW; F4954E7DC70FD233 CRC64;

Query Match 4.38; Score 102.5; DB 1; Length 654;
Best Local Similarity 24.08; Pred. No. 4.2; Indels 103; Gaps 18;
Matches 78; Conservative 35; Mismatches 109;

QY 182 CFSNFIASKL-----EAVEYAHISRLQADL--NCLSDL-----LRSSTQWKYIN 225
Db 377 CFGGVVAEAVGIDKDKREAVTVA-VKMLDPAATEKDLSDLVSEMMKMKIKHKNIN 435
QY 226 L---CGGFPIKSNFELVSELKINGANMLETVPKPSKLEFTHHELRVPEYVKLP 282
Db 436 LIGACTQDGPL--YVIVEYASKGNLREYLRARRPQME-----YSTDKNVPEQMTF- 486
QY 283 IRTNISKAPPHNIQIFVGSAYFY-----LSQAFVKYIFENNSIVQDFPWSKDTYSP 334
Db 487 -----KLVSGTYQLANGMEYIASQKCIHRDLARNVLYENNVMKIADEGLARDINNI 540
QY 335 DEHFATLIRVPGIRGELISRSQDVSDLOSKTRLVKNYVYEGFFPSCGSHLSRYCIY 394
Db 541 DYKKTTRGRLP-----VKWMAFPALDLYVT--HQSDVWSFG 576
QY 395 -----AELRLIKDGHWFANKFDSKVDPIILIKIAELKEQQRD-W 435
Db 577 VLMEIFTLGSGPYGIVVELEFKLKRGH---RMDKPN-----CTNE-LYMMMDCW 626
QY 436 IILPSEK-LF-----MDRNLTTTS 453
Db 627 HAVPSQRPFTKQVLEDRLITLIT 651

RESULT 9
BFR2_HUMAN
ID BFR2_HUMAN STANDARD; PRT; 821 AA.
AC P21802; P18443;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR 2 PRECURSOR (FGFR-2) (EC 2.7.1.112)
DE (KERATINOCYTE GROWTH FACTOR RECEPTOR).
GN FGFR2 OR BEK OR BFR-1 OR KSAM-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_Taxid=9606;
RX MEDLINE-91045561; PubMed-2172978;
RX MEDLINE-90360977; PubMed-1697263;
RX Diome C.A., Crumley G.R., Bellot F., Kaplow J.M., Searfoss G.,
RX Ruta M., Burgess W.H., Jaye M., Schlessinger J.;
RX "Cloning and expression of two distinct high-affinity receptors
RX cross-reacting with acidic and basic fibroblast growth factors.";
RL EMBO J. 9:2685-2692(1990).
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Neonatal brain stem;
RX MEDLINE-91045561; PubMed-2172978;
RA Houssaint E., Blanquet P.R., Champion-Arnaud P., Gesnel M.C.,
RA Torriglia A., Courtois Y., Breathnach R.;
RA "Related fibroblast growth factor receptor genes exist in the human
RA genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8180-8184(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-91274356; PubMed-1647213;
RA Seno M., Sasada R., Watanabe T., Ishimaru K., Igarashi K.;

```

RT "Two cDNAs encoding novel human EGF receptor.";  
 RL Biochim. Biophys. Acta 1089:244-246(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Stomach cancer;  
 RX MEDLINE-90332706; PubMed-2377625.  
 RA Hattori Y., Odagiri H., Nakatani H., Miyagawa K., Naito K.,  
 RA Sakamoto H., Katoh O., Yoshida T., Sugimura T., Terada M.,  
 RT "K-sam, an amplified gene in stomach cancer, is a member of the  
 RT heparin-binding growth factor receptor genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5983-5987(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92212948; PubMed-1313574;  
 RA Katoh M., Hattori Y., Sasaki H., Tanaka M., Sugano K., Yazaki Y.,  
 RA Sugimura T., Terada M.;  
 RT "K-sam gene encodes secreted as well as transmembrane receptor  
 RT tyrosine kinase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2960-2964(1992).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RX MEDLINE-93016048; PubMed-1400433;  
 RA Dell K.R., Williams L.T.;  
 RT "A novel form of fibroblast growth factor receptor 2. Alternative  
 RT splicing of the third immunoglobulin-like domain confers ligand  
 RT binding specificity.";  
 RL J. Biol. Chem. 267:21225-21229(1992).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Blood;  
 RA Steinberger D., Mueller U.;  
 RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP VARIANTS CS HIS-340, ARG-342, SER-342, TYR-342 AND CYS-354.  
 RX MEDLINE-95078932; PubMed-7987400;  
 RA Reardon W., Winter R.M., Rutland P., Pulley L.J., Jones B.M.,  
 RA Malcolm S.;  
 RT "Mutations in the fibroblast growth factor receptor 2 gene cause  
 RT Cronzon syndrome.";  
 RL Nat. Genet. 8:98-103(1994).  
 RN [9]  
 RP VARIANTS CS CYS-328 AND CYS-347, AND VARIANT JWS GLY-344.  
 RX MEDLINE-95179174; PubMed-874170;  
 RA Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J.,  
 RA Charnas L.R., Jackson C.E., Jaye M.;  
 RT "Jackson-Weiss and Cronzon syndromes are allelic with mutations in  
 RT fibroblast growth factor receptor 2.";  
 RL Nat. Genet. 8:275-279(1994).  
 RN [10]  
 RP VARIANTS CS.  
 RX MEDLINE-95384152; PubMed-7655462;  
 RA Rutland P., Wilkie A.O.M., Slaney S.F., Poole M.D., Pulley L.J.,  
 RA Rutland P., Hockley A.D., Wake M.J.C., Goldin J.H., Winter R.M.,  
 RA Reardon W., Malcolm S.;  
 RT "Mutations in the third immunoglobulin domain of the fibroblast growth  
 RT factor receptor 2 gene in Cronzon syndrome.";  
 RL Hum. Mol. Genet. 4:1077-1082(1995).  
 RN [11]  
 RP VARIANT PS ALA-321.  
 RX MEDLINE-95235551; PubMed-7719333;  
 RA Lajeune E., Wei M.H., Bonaventure J., Munnich A., Le Merrer M.,  
 RA Renier D.;  
 RT "FGFR2 mutations in Pfeiffer syndrome.";  
 RL Nat. Genet. 9:108-108(1995).  
 RN [12]  
 RP VARIANTS AS TRP-252 AND ARG-253.  
 RX MEDLINE-95235562; PubMed-7719344;  
 RA Wilkie A.O.M., Slaney S.F., Oldridge M., Poole M.D., Ashworth G.J.,  
 RA Hockley A.D., Hayward R.D., David D.J., Pulley L.J., Rutland P.,  
 RA Malcolm S., Winter R.M., Reardon W.,  
 RT "Apert syndrome results from localized mutations of FGFR2 and is  
 RT allelic with Cronzon syndrome.";  
 RL Nat. Genet. 9:165-172(1995).  
 RN [13]  
 RP VARIANTS PS PRO-341, ARG-342 AND TYR-342.  
 RX MEDLINE-95235563; PubMed-7719345;  
 RA Rutland P., Pulley L.J., Reardon W., Baraister M., Hayward R.,  
 RA Jones B., Malcolm S., Winter R.M., Oldridge M., Slaney S.F.,  
 RA Poole M.D., Wilkie A.O.M.;  
 RT "Identical mutations in the FGFR2 gene cause both Pfeiffer and  
 RT Cronzon syndrome phenotypes.";  
 RL Nat. Genet. 9:173-176(1995).  
 RN [14]  
 RP VARIANTS CS/JWS/PS.  
 RX MEDLINE-96203627; PubMed-8644708;  
 RA Meyers G.A., Day D., Goldberg R., Daentl D.L., Przylerka K.A.,  
 RA Abrams L.J., Graham J.M. Jr., Felngold M., Moeschler J.B.,  
 RA Ramsley E., Scott A.F., Jabs E.W.;  
 RT "FGFR2 exon IIIa and IIc mutations in Cronzon, Jackson-Weiss, and  
 RT Pfeiffer syndromes: evidence for missense changes, insertions, and a  
 RT deletion due to alternative RNA splicing.";  
 RL Am. J. Hum. Genet. 58:491-498(1996).  
 RN [15]  
 RP VARIANTS CS CYS-105, GLU-338, CYS-351 AND ARG-384.  
 RX MEDLINE-97101656; PubMed-8946174;  
 RA Pulley L.J., Reardon W., Wilkes D., Rutland P., Jones B.M.,  
 RA Hayward R., Hall C.M., Brueton L., Chun N., Lammer E., Malcolm S.,  
 RA Winter R.M.;  
 RT "Spectrum of craniosynostosis phenotypes associated with novel  
 RT mutations at the fibroblast growth factor receptor 2 locus.";  
 RL Eur. J. Hum. Genet. 4:283-291(1996).  
 RN [16]  
 RP VARIANTS CS ILE-331, ASP-336--ALA-337 DUPLE AND TRP-356--THR-358 DEL.  
 RX MEDLINE-97114301; PubMed-8956050;  
 RA Steinberger D., Mulliken J.B., Mueller U.;  
 RT "Cronzon syndrome: previously unrecognized deletion, duplication, and  
 RT point mutation within FGFR2 gene.";  
 RL Hum. Mutat. 8:386-390(1996).  
 RN [17]  
 RP VARIANT PS CYS-290.  
 RX MEDLINE-97295073; PubMed-9150725;  
 RA Tartaglia M., Valeri S., Valardi F., di Rocco C., Battaglia P.A.;  
 RT "Trp90Cys mutation in exon IIIa of the fibroblast growth factor  
 RT receptor 2 (FGFR2) gene is associated with Pfeiffer syndrome.";  
 RL Hum. Genet. 99:602-606(1997).  
 RN [18]  
 RP VARIANT CS L-252, VARIANT AS F-252, AND VARIANT PS F-252--S-253.  
 RX MEDLINE-97156222; PubMed-9002682;  
 RA Oldridge M., Lunt P.W., Zackai E.H., McDonald-Mogil D.M., Muenke M.,  
 RA Moloney D.M., Twigg S.R.F., Heath J.K., Howard T.D., Hoganson G.,  
 RA Gagnon D.M., Jabs E.W., Wilkie A.O.M.;  
 RT "Genotype-phenotype correlation for nucleotide substitutions in the  
 RT Igit-Igit1 linker of FGFR2.";  
 RL Hum. Mol. Genet. 6:137-143(1997).  
 RN [19]  
 RP VARIANT CS GLU-292.  
 RX MEDLINE-97297373; PubMed-9152842;  
 RA Steinberger D., Collmann H., Schmalenberger B., Mueller U.;  
 RT "A novel mutation (a886g) in exon 5 of FGFR2 in members of a family  
 RT with Cronzon phenotype and plagiocephaly.";  
 RL J. Med. Genet. 34:420-422(1997).  
 RN [20]  
 RP VARIANTS CS VAL-276; CYS-301 AND SER-314.  
 RX MEDLINE-98180879; PubMed-9521581;  
 RA Steinberger D., Vriend G., Mulliken J.B., Mueller U.;  
 RT "The mutations in FGFR2-associated craniosynostoses are clustered in  
 RT five structural elements of immunoglobulin-like domain III of the  
 RT receptor.";  
 RL Hum. Genet. 102:145-150(1998).  
 RN [21]  
 RP VARIANTS AS TRP-252 AND ARG-253.  
 RX MEDLINE-98112406; PubMed-9452027;  
 RA Tsai F.-J., Hwu W.-L., Lin S.-P., Chang J.-G., Wang T.-R., Tsai C.-H.;  
 RT "Two common mutations 934C to G and 937C to G of fibroblast growth  
 RT factor receptor 2 (FGFR2) gene in Chinese patients with Apert



```

FT CAROHND 147. 147 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHND 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHND 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHND 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHND 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHND 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHND 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARSPLC 37 37 E -> G (IN SHORT ISOFORM)
FT VARSPLC 152 152 MISSING (IN SHORT ISOFORM)
FT VARSPLC 361 361 MISSING (IN SHORT ISOFORM)
FT VARSPLC 361 361 MISSING (IN SHORT ISOFORM)
FT CONFLICT 169 169 AMYLVL -> HSGINSSNAEVLALFNTEDAGEYICKVS
FT CONFLICT 187 187 NYIGQANSAMLTVPKQ (IN SHORT ISOFORM)
SQ SEQUENCE 821 AA; 91983 MW; FQDB28ADD61F4414 CRC64;

Query Match 4.3%; Score 102.5; DB 1; Length 821;
Best Local Similarity 24.0%; Pred. No. 5.7;
Matches 78; Conservative 35; Mismatches 109; Indels 103; Gaps 18;

QY 182 CFSNIFIAKL-----EAVEVAHISRLQADL--NCLSDL-----LKSSIQMKRYIN 225
DB 491 CGQYVMAAAGVGDKDKPKENAVTA-VKMLKQDATEKDSLDVSEMKMKIGKRNIN 549
QY 226 L--CGQDFPLASNFELVSELKLCNANLFTVKKPNSKLEFTHHEIRRVPEYVKLP 282
DB 550 LIGACTQDPL--YVIVYASGKNLEKRLRARPGE--YSDINRPEEQMTF- 600
QY 283 IRTNISKEAPPHNIOIFVASAYV-----LSQAFYVIFNNSIVQDFMKSQTSF 334
DB 601 -----KDLVSCYQIARMEYIASOKCHRDLAARNVLTENNWKIADFGIARDINN 654
QY 335 DEHFMAITLRVPGIPGEISRSADVSDLSQKTRLVKWNYYEGEFPYSGSHLSVCYIG 394
DB 655 DYKKTNGRLP-----YKMAPEALFDRVY--HQSDVWSPG 690
QY 395 -----AAELRLIKDGHWFANKRDSVDPILITGLAEKLEBQRD-W 435
DB 691 VLMFEITLGGSPYPGIPVELEFKLKEGH---RMDKPTN---CTNE-LYMMRDCW 740
QY 436 ITLPSEK-LF-----MDRLTTTS 453
DB 741 HAVPSQAPTEKQVLEDRLITLT 765

RESULT 11
YCF2_PINTH STANDARD; PRT; 2054 AA.
AC P1653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 244.6 KDa PROTEIN (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugita M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC - FUNCTION: NOT YET KNOWN.
CC - SIMILARITY: BELONGS THE YCF2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D17510; BAA04460.1; -
DR InterPro: IPR001939; -
DR Pfam: PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2054;
Best Local Similarity 21.5%; Pred. No. 23;
Matches 72; Conservative 60; Mismatches 98; Indels 105; Gaps 20;

QY 15 KVFLLFTLW-LLSLKLINVRLEFPQKDIYVEY-SLSTSPFVRNRYTHVKDEYREVN 72
DB 864 QTFVLIDYLVRSFNLRLRP--FVREKRLSSIEISTPILTKEDI-----VNEKN 914
QY 73 -CSGYDEPPEIKSLERRDITLEDDEYVAMTSQCDIYQTLRGAQKLYS-----K 126
DB 915 FCGPFFKSDSENNPQCFKR-----GSSNVGLIQ--RKYQDOLLSEMSNK 962
QY 127 EKESEPIAYSIVHKDALMELRLHAIFYNQHNITCIHDKRAPTEFVAMNNAKCSNI 166
DB 963 NEEIFPR--QDVEFTECLKNKTYNE-----DIDGS-----TSSNSKEBONT 1004
QY 187 FIASKLEVEVAHISRLQADLNCLSLKSISQW-----KYVINCQDFPLKSNFEL 239
DB 1005 YRISQIDSI--FSKW-----DLFKTYMPFFTSAMKYLENNL--LDT 1043
QY 240 VSELKLCNAN--METVAPNSKLEFTHHEIRRVPEYVKLPDIRNISKEAPPHN 295
DB 1044 LSEI-LHGSNPFSIIONIK-HNILLKR-NILMELSHPLMEPIQCKLRNLTIN----- 1094
QY 296 IQIFGSAVFLSQAFVYIFNNSIVQDFMWSMD 330
DB 1095 -----KFFPSNNKDFEYCKD 1112

RESULT 12
YCF2_TOBAC STANDARD; PRT; 2280 AA.
AC P09976; P09977;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEICAL 267 KDa PROTEIN (ORF 2280).
GN YCF2.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanales; Solanales; Solanales; Solanales; Solanales;
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, BRIGHT YELLOW 4;
RA Sugita M.;
RL Submitted (AUG-1986) to the EMBL/GenBank/DBJ databases.
RN [2]
RP COMPLETE GENOME.
RA Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohno C., Torizawa K., Meng B.Y., Sugita M.,
RA Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA Tohoh N., Shimada H., Sugita M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
RT its gene organization and expression.";
RL EMBO J. 5:2043-2049(1986).
RN [3]

```





DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
DR PROSITE; PS00111; PROTEIN KINASE\_DOM; 1.  
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;  
KW transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;  
KW Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 823 TYROSINE KINASE RECEPTOR CEK3.  
FT DOMAIN 24 379 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 380 400 POTENTIAL.  
FT DOMAIN 401 823 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 58 117 IG-LIKE DOMAIN.  
FT DOMAIN 174 240 IG-LIKE DOMAIN.  
FT DOMAIN 273 351 IG-LIKE DOMAIN.  
FT DOMAIN 135 145 ASP/GLU-RICH (HIGHLY ACIDIC).  
FT DOMAIN 483 772 ASB/GLU-RICH (HIGHLY ACIDIC).  
FT NP\_BIND 489 497 PROTEIN KINASE.  
FT BINDING 519 519 ATP (BY SIMILARITY).  
FT ACT\_SITE 628 628 ATP (BY SIMILARITY).  
FT MOD\_RES 659 659 BY SIMILARITY.  
FT DISULFID 65 110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT DISULFID 181 233 POTENTIAL.  
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 823 AA; 92299 MW; 42BF3C64EA02FD43 CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 823;  
Best Local Similarity 24.0%; Pred. No. 9.4;  
Matches 78; Conservative 34; Mismatches 110; Indels 103; Gaps 18;

QY 182 CFSNFIASKL-----EAVEAHISRLQADL--NCISDL-----LKSSIQMKYIN 225  
DB 493 CFGGVNAEAVGIDKRPKEAVYA-VKMLKDATEKDISDLVSEMEKMKIGKHNIIN 551  
QY 226 L---CGQDFPLKSNFELVSEIKINGANMLETVKPPNSKLERFTYHHELRVPYEVKLP 282  
DB 552 LLGACTQDDGL--YVIVEYASKNLEREYLARPPGME----YSPDINRVPEQMTE- 602  
QY 283 IRTNISKAPPHNIQIVGSAHYV-----LSQAFVKIIFNNSTIVQDFFANSKDTYSP 334  
DB 603 -----KDLVSCYQLARGMEYLASQKCIHRDLAARNVLTENNVMKIADEGLARDINNI 656  
QY 335 DEHFMAFLIRPGIPGEISRSADVDLSQKTRLVKWNVEGFYFPGCTGSHLSVCITYG 394  
DB 657 DYVKKTNGRLP-----VKMAPEALEFDRYIT--HOSDWMSFG 692  
QY 395 -----AAELRWLIKDGHWFANKEDSKVDPILIKCLAEKLEEQORD-W 435  
DB 693 VLWMEIFTLGSSPYPGIPVELEFKLEKGH-----RMDKPPAN-----CTNE-LYMMHDCW 742  
QY 436 ITPSEK-LF-----MDRNLTTTS 453  
DB 743 QAVPSQRPTRFQVLVEDLDRLITLT 767

Search completed: August 2, 2001, 14:50:18  
Job time: 298 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2001, 00:11:58 ; Search time 113.87 Seconds  
(without alignments)  
7510.336 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362

Sequence: 1 atgaagataatcaatgttta.....atctactacacacatcatga 1362

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	22 AAS00045	Human DNA encoding
2	191.8	14.1	1221	21 AAA96575	Core 2 beta-1,6-N-
3	191.8	14.1	1317	21 AAA96569	DNA encoding a cor
4	191.8	14.1	2108	21 AAA96570	A core 2 beta-1,6-
5	191.8	14.1	2128	22 AAC62134	Nucleotide sequenc
6	191.8	14.1	2229	21 AAC98109	Human pancreatic c
7	191.8	14.1	2319	21 AAA48623	Human C2/4GNT CDNA
8	185.2	13.6	1807	16 AAQ89201	I-branchning enzyme
9	185.2	13.6	1807	17 AAT16201	Beta-1,6-N-acetylgl
10	185.2	13.6	1807	19 AAV30006	Full length cDNA s
11	185.2	13.6	1807	19 AAV16000	Human beta-1,6-N-a

12	183	13.4	186	21 AAC19209	Human secreted pro
13	175.2	12.9	2105	15 AAQ61559	CDNA sequence of h
14	174.4	12.8	5010	19 AA24042	Rat DH1 CDNA. Rat
15	174.4	12.8	5010	21 AAC65468	Diabetic rat heart
16	172	12.6	2204	19 AA24043	Human core 2GNT DN
17	172	12.6	2204	21 AAC65469	Human heart core 2
18	160.8	11.8	2102	16 AAT80112	Coding sequence fo
19	121.2	8.9	378	19 AAV30005	CDNA encoding a re
20	112.4	8.3	936	22 AAF58252	Oligonucleotide D1
21	112.4	8.3	936	22 AAF58254	Oligonucleotide D1
22	112.4	8.3	936	22 AAF58257	Oligonucleotide D1
23	112.4	8.3	936	22 AAF58259	Oligonucleotide D1
24	112.4	8.3	936	22 AAF58262	Oligonucleotide D2
25	112.4	8.3	936	22 AAF58265	Oligonucleotide D2
26	110.6	8.1	936	22 AAF58252	Oligonucleotide D1
27	110.6	8.1	936	22 AAF58254	Oligonucleotide D1
28	110.6	8.1	936	22 AAF58257	Oligonucleotide D1
29	110.6	8.1	936	22 AAF58259	Oligonucleotide D1
30	110.6	8.1	936	22 AAF58262	Oligonucleotide D2
31	110.6	8.1	938	22 AAF58255	Oligonucleotide D1
32	92.2	5.8	378	17 AAT16202	C2GNT catalytic do
33	80.8	5.9	997	19 AAF59800	Human secreted pro
34	66.2	4.9	300	21 AA401261	Human colon cancer
35	53.4	3.9	192	22 AAC62139	Murine beta-1,6-N-
36	49.4	3.6	777	19 AAF59681	Human secreted pro
37	47.4	3.5	307	21 AAC01854	Human secreted pro
38	46.2	3.4	6171	19 AAF52170	Streptococcus pneu
39	41.8	3.1	244	22 AAF58238	Oligonucleotide D1
40	40.6	3.0	244	22 AAF58238	Oligonucleotide D1
41	40.6	3.0	7430	19 AAF31250	E. coli y96 pathog
42	35.8	2.6	814	20 AAX22268	Human secreted pro
43	35.2	2.6	99	17 AAT16203	C2GNT C-terminal r
44	35.2	2.6	322	21 AAT79284	Human lung tumour
45	35	2.6	783	20 AAT16673	Human gene express

#### ALIGNMENTS

RESULT 1	
AAS00045	standard; DNA: 1362 BP.
XX	
AC	AAS00045;
XX	
DT	14-MAY-2001 (first entry)
XX	
DE	Human DNA encoding C2GNT3.
XX	
KW	Human; C2GNT3; Thymus-related disorder; cancer; tumour; adenoma;
KW	UPP-GICNAC; Galbeta1,3GalNAc alpha-R beta1,6GICNAC transferase; sarcoma;
KW	malignant melanoma; breast cancer; cervical cancer; hypocoactivity;
KW	hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KW	leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KW	acquired immunodeficiency syndrome; sepsis; wound healing; infection; ds.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	CDS
FT	
FT	1..1362
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= "C2GNT3"
FT	complement (114..138)
FT	/*tag= b
FT	/label= "Binding site for PCR primer TSHC100"
FT	1367..1362
FT	/*tag= c
FT	/label= "Binding site for PCR primer TSHC121"
PN	WO200114535-A2.
XX	
PD	01-MAR-2001.
XX	



PE 24-AUG-2000; 2000WC-DK00469.  
 XX 24-AUG-1999; 99US-0150488.  
 PR (SCHW/) SCHWIENTEK T.  
 PA (CLAU/) CLAUSEN H.  
 XX Schlientek T, Clausen H;  
 PI WPI: 2001-228615/23.  
 DR P-FSDB: AAU00037.  
 XX  
 XX New C2GnT3 polypeptides and nucleic acids encoding the polypeptides  
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,  
 PT thymus-related disorders, cancers, tumours, immunosuppression  
 XX  
 PS Claim 5; Fig 1; 97pp; English.

The sequence encodes human UDP-N-acetyl-glucosamine:Galactose- $\beta$ -galactose- $\beta$ -N-acetylglucosaminyltransferase (UDP-GlcNAc:Gal $\beta$ Gal $\beta$ 6-GlcNAc transferase or C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the preparation of compositions for treating conditions mediated by C2GnT3, particularly a thymus-related disorder. C2GnT3, nucleic acids encoding it and antibodies against it may also be used for in vitro purposes related to scientific research, DNA synthesis and manufacture of vectors, in the prognostic and diagnostic evaluation of conditions associated with altered expression or activity of C2GnT3 or conditions requiring modulation of C2GnT3, as well as in monitoring conditions by detecting and localising the DNA and protein. Disorders such as tumours (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the breast or cervix), hypoparathyroidism, hyperactivity, atrophy, enlargement of thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression, acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome, sepsis, wound healing, acute and chronic infection, cell-mediated or humoral immunity, or TNF/TN2 imbalance, may be treated using these protein or nucleic acid. The antibodies may be used to screen potential therapeutic compounds to determine their effects on a conditions such as thymus-related disorder or cancer, to determine the level of C2GnT3 expression in cells genetically engineered to produce C2GnT3, or to detect and quantify polypeptides in a sample to determine their role in a particular cellular events or pathological states and to diagnose and treat such pathological states.

Sequence 1362 BP; 410 A; 260 C; 279 G; 413 T; 0 other;

Query Match 100.0%; Score 1362; DB 22; Length 1362;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatattcaaatgtattttaacatacctacagcagaagtttcatcctttt 60  
 Db 1 atgaagatattcaaatgtattttaacatacctacagcagaagtttcatcctttt 60  
 QY 61 ttaacccatgtcgtcctcctttttaaagcttctaagtgaagacattcttcgcaa 120  
 Db 61 ttaacccatgtcgtcctcctttttaaagcttctaagtgaagacattcttcgcaa 120  
 QY 121 aaagacattacttggtgagtaaccttaagtaacctgaccttttgaagaagaagttac 180  
 Db 121 aaagacattacttggtgagtaaccttaagtaacctgaccttttgaagaagaagttac 180  
 QY 181 actcatgttaagatgaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 240  
 Db 181 actcatgttaagatgaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 240  
 QY 241 cctttggaattggaagaagctctggaataagaagaaggaaggaaggaaggaaggaagga 300  
 Db 241 cctttggaattggaagaagctctggaataagaagaaggaaggaaggaaggaaggaagga 300  
 QY 301 gatgatgttggaagaatgacagatgattgacattatcagactctaagaagttatgct 360  
 Db 301 gatgatgttggaagaatgacagatgattgacattatcagactctaagaagttatgct 360

Db 301 gatgatgttggaagaatgacagatgattgacattatcagactctaagaagttatgct 360  
 QY 361 caaaagcttgctccaagaagaggaagaagcttcccaatgactattcttggttccac 420  
 Db 361 caaaagcttgctccaagaagaggaagaagcttcccaatgactattcttggttccac 420  
 QY 421 aaagatgcaattatgtgttaaaagccttccatgcatatatacaacagacatatattac 480  
 Db 421 aaagatgcaattatgtgttaaaagccttccatgcatatatacaacagacatatattac 480  
 QY 481 tgcattcatatgatactgtaaggaacactgaaccttaacgttgcaatgacaattagct 540  
 Db 481 tgcattcatatgatactgtaaggaacactgaaccttaacgttgcaatgacaattagct 540  
 QY 541 aagtgcttcccaatttttccattgcttcccaattgagagcgtctggaatcccaatt 600  
 Db 541 aagtgcttcccaatttttccattgcttcccaattgagagcgtctggaatcccaatt 600  
 QY 541 aagtgcttcccaatttttccattgcttcccaattgagagcgtctggaatcccaatt 600  
 Db 541 aagtgcttcccaatttttccattgcttcccaattgagagcgtctggaatcccaatt 600  
 QY 601 tccagatccagctgatttaatttgcttggaaccttctgaagtcctcaatccagttg 660  
 Db 601 tccagatccagctgatttaatttgcttggaaccttctgaagtcctcaatccagttg 660  
 QY 661 aatatgcttaccattgtgttggaagatttccctgaagtcgaattttgaatttggtg 720  
 Db 661 aatatgcttaccattgtgttggaagatttccctgaagtcgaattttgaatttggtg 720  
 QY 721 tccagatggaagaacatccatggaagaatattgttggaagcgttgaaaccccaacagt 780  
 Db 721 tccagatggaagaacatccatggaagaatattgttggaagcgttgaaaccccaacagt 780  
 QY 781 aatttgaagaatgcttaccatgcatatgacttagagaggttccctatgaaatattgag 840  
 Db 781 aatttgaagaatgcttaccatgcatatgacttagagaggttccctatgaaatattgag 840  
 QY 841 ctaccataaggaacaacatctccaaaggaagaccccccaataacattcaatatttgt 900  
 Db 841 ctaccataaggaacaacatctccaaaggaagaccccccaataacattcaatatttgt 900  
 QY 901 ggaagcttattttgttttaagtaagaagattgttaaatattttcaaacaccatc 960  
 Db 901 ggaagcttattttgttttaagtaagaagattgttaaatattttcaaacaccatc 960  
 QY 961 gtccaagactttttgctggttcaagaacacatactctccgaatgagactttggct 1020  
 Db 961 gtccaagactttttgctggttcaagaacacatactctccgaatgagactttggct 1020  
 QY 1021 acctgattcgggttccagaataacctggtggaagatttccagatccagccaagatgtgt 1080  
 Db 1021 acctgattcgggttccagaataacctggtggaagatttccagatccagccaagatgtgt 1080  
 QY 1081 gatctgcaagtaagaactgacctgtccaagtgaattactatgaagagcttttccatcc 1140  
 Db 1081 gatctgcaagtaagaactgacctgtccaagtgaattactatgaagagcttttccatcc 1140  
 QY 1141 agttgactggaactcactccttgaagcgtgtgtattttagggcgtgcaagaattaaagttg 1200  
 Db 1141 agttgactggaactcactccttgaagcgtgtgtattttagggcgtgcaagaattaaagttg 1200  
 QY 1201 ctatcaagaatgacatgattgttgtaataatttgatttcaagtgtgacctatctgt 1260  
 Db 1201 ctatcaagaatgacatgattgttgtaataatttgatttcaagtgtgacctatctgt 1260  
 QY 1261 attaatgcttgagcaagaagcttgagaagaacagagagactgatacatttgcctca 1320  
 Db 1261 attaatgcttgagcaagaagcttgagaagaacagagagactgatacatttgcctca 1320  
 QY 1321 gaaagttatttatgatagaatactcactacacatactga 1362  
 Db 1321 gaaagttatttatgatagaatactcactacacatactga 1362

RESULT 2  
 AAA96575

ID AAA96575 standard; DNA; 1221 BP.  
 XX  
 AC AAA96575;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.  
 XX  
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;  
 XX cancer: cardiovascular disorder; inflammatory disorder; asthma;  
 XX rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;  
 XX septic shock; adult respiratory distress syndrome; ARDS; cancer;  
 XX platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;  
 XX clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;  
 XX diverticulitis; ulcerative colitis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX CA226936-A1.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 03-FEB-2000; 2000CA-2296936.  
 XX  
 PR 03-FEB-1999; 99US-0118674.  
 XX  
 PA (GLYC-) GLYCDESIGN INC.  
 XX  
 PI Korczak B, Lew A;  
 XX  
 DR WPI; 2000-594746/57.  
 XX  
 PT New nucleic acid molecules of core 2  
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new  
 PT compositions for treatment of disorders mediated by the enzyme  
 PT including cancer, cardiovascular and inflammatory disorders.  
 XX  
 PS Claim 4; Page 54; 66pp; English.  
 XX  
 CC The present sequence encodes a fragment of a human core 2  
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.  
 CC The polypeptide can be used to treat diseases and disorders, such as  
 CC cancer, cardiovascular disorders and inflammatory disorders including  
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,  
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome  
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as  
 CC atherosclerosis and clotting can also be treated. The polypeptides of  
 CC the invention are predominantly expressed in gastrointestinal tissue  
 CC (stomach, colon, intestine, testis) and are elevated in cancer.  
 CC Gastrointestinal disorders that may be prevented or treated include  
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis  
 CC and ulcerative colitis. The antibodies may be used in  
 CC immuno-histochemical analysis, to detect the novel polypeptide and to  
 CC localize it to particular cells and tissues and to specific subcellular  
 CC locations and to quantitate the level of expression.  
 XX  
 SQ Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 Other;

Query Match 14.1%; Score 191.8; DB 21; Length 1221;  
 Best Local Similarity 52.5%; Pred. No. 4.2e-44;  
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

Db 341 ttgaagagctacgagcgtgtatgtccctcagaacatactgtgtccatgtgagatg 400  
 Qy 497 gtaagcaccctgatactcctcaagaattgacatgaacaaattagctaagtcttccaaata 556  
 Db 401 agaagtcaccagaaccttccaagaagcggtccaagaacattattctgttccccaatg 460  
 Qy 557 ttctatgtctccaattgagagctgtggaatactgcccacatttccagactcagaatg 616  
 Db 461 tcttcataagcagaatagctgttgcgtgtttatgtcttccctcgtgtccaggtgtcaagct 520  
 Qy 617 atttaattgtcttgcgaccttctgaagttctcaatcagatggaataatgttcaact 676  
 Db 521 accccaactgatatgaagacttgcctcagagctcagcgctgtgaataactcttgata 580  
 Qy 677 tgtgtggcaagatttccctcctgaagcacaatttgaattgtgttcaagattgaaaac 736  
 Db 581 catgtggagcagacttctcctataaagagcagatgcagagatggtccagctcctaagtgt 640  
 Qy 737 tcaatgagcaaatatgttgagacggttgaaccccccaacagtaatttggaaagtca 796  
 Db 641 tgaatggagaggaatagcatgtgacagagatcctcctaagcacaagaacccgctgga 700  
 Qy 797 ctaccatcatgaacttgaacggtgtccttatgtatatgtgaagtccaataagagca 856  
 Db 701 aatatcacttggagtagtgagagacaataac-----ctaaca 742  
 Qy 857 acatcccaaggaagcaccoccccaatacattatgttggcagtgctatttg 916  
 Db 743 acaagaagaagatctctccctcctataatttaactatgttcaagaggaatgcgtacatg 802  
 Qy 917 tttaagtcgaagcattgttaataatatattccaacaacatccatcgttccaagacttttg 976  
 Db 803 tggcttcccgagatttgcacaacatgttgaagaacctcaaaatccacaactgattg 862  
 Qy 977 cctgttcaagacacatacctcctcgtgaagcactttgtggtcacttgaatggttc 1036  
 Db 863 aatgggtaaagacacctatagccacagatgaacacctgtggccaccttcagcggtcac 922  
 Qy 1037 caggaataccttggagagatttccagatcagcccaagatg---tgtcatctgcagata 1093  
 Db 923 gttgagatgctgtcctcttcccaaccacccaagatcagacatccagatgacttcta 982  
 Qy 1094 agaccgcagcttcaagaggaattactatgaaggtttctatccagct----- 1143  
 Db 983 ttgcagagctgttcaagtgccgaggtctatgtaggaggaatcagataaggtgtcctta 1042  
 Qy 1144 -----tgtctgatactcaacttcgaagcgtgtgattatgaagctcagaataagt 1198  
 Db 1043 ctccctgcttgaatccaccacagcggtatctgttattgtggtggagctgaatt 1102  
 Qy 1199 ggcttatcaagaatgagacttgttgcataaattgatctagaagtgcacctact 1258  
 Db 1103 ggaatgttaaaacatccaccctgttgcacaagaatttgaaccaagtgagtgataatg 1162  
 Qy 1259 tgatnaatgtcttgcagaagaagct 1283  
 Db 1163 ccttcagtgcttaagaagataact 1187

RESULT 3  
 AAA96569  
 ID AAA96569 standard; DNA; 1317 BP.  
 XX  
 AC AAA96569;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.  
 XX  
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;  
 XX cancer: cardiovascular disorder; inflammatory disorder; asthma;  
 XX rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;  
 XX septic shock; adult respiratory distress syndrome; ARDS; cancer;

KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;  
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;  
 KW diverticulitis; ulcerative colitis; ss.

OS Homo sapiens.

Key Location/Qualifiers  
 CDS 1..1317

FT /\*tag= a  
 FT /product= "core 2  
 FT Beta-1,6-N-acetylglucosaminyltransferase"

XX CA2296936-A1.

XX 03-AUG-2000.

XX 03-FEB-2000; 2000CA-2296936.

XX 03-FEB-1999; 99US-0118674.

XX (GLYC-) GLYCDESIGN INC.

XX Korczak B, Lew A;

XX WPI: 2000-594746/57.

XX P-PSDB; AAB18995.

XX New nucleic acid molecules of core 2

XX beta-1,6-N-acetylglucosaminyltransferase useful for providing new

XX compositions for treatment of disorders mediated by the enzyme

XX including cancer, cardiovascular and inflammatory disorders.

XX Claim 4; Page 50; 66pp; English.

XX The present sequence encodes a human core 2

XX beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.

XX The polypeptide can be used to treat diseases and disorders, such as

XX cancer, cardiovascular disorders and inflammatory disorders including

XX asthma, rheumatoid arthritis, inflammatory bowel disease,

XX arteriosclerosis, septic shock, adult respiratory distress syndrome

XX (ARDS) and cancer. Various platelet-mediated pathologies such as

XX atherosclerosis and clotting can also be treated. The polypeptides of

XX the invention are predominantly expressed in gastrointestinal tissue

XX (stomach, colon, intestine, testis) and are elevated in cancer.

XX Gastrointestinal disorders that may be prevented or treated include

XX ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis

XX and ulcerative colitis. The antibodies may be used in

XX immuno-histochemical analysis, to detect the novel polypeptide and to

XX localize it to particular cells and tissues and to specific subcellular

XX locations and to quantitate the level of expression.

XX Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;

Query Match 14.1%; Score 191.8; DB: 21; Length 1317;  
 Best Local Similarity 52.5%; Pred. No. 4.3e-44;  
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagttattgtacattatcagactcagaggttatgtcacaagaattgtcaca 376  
 DB 317 taccagagactgtgagcttcaagctggaaggaatttatagatcccaactgagca 376  
 QY 377 aggaagagaagaagcttcccaatagcctattcttggtgtccacaaagatccaattatg 436  
 DB 377 aagaagaggttgagttccctattgcatctatgttattatgagaaattggaact 436  
 QY 437 ttgaagagtttcatgtatatacaaccagacacattatcagctcattatgac 496  
 DB 437 ttgaagagtttcatgtatatacaaccagacacattatcagctcattatgac 496  
 QY 497 gtaagagcattatataaagtgtcacaatgaacaattttagtaagtgtctccata 556  
 DB 497 agaagtcctccagaacttcaagaagcggttcaagaactatttctgtccccaatg 556

RESULT 4  
 AAA96570  
 ID AAA96570 standard; DNA; 2108 BP.

QY 557 ttctcatgctcccaattagagctgtggaatatgagccacatttccagactcagctg 616  
 DB 557 tctctatagcagtaagctggttccgtgttattatgctctccgtccaggtgtgcaagctg 616  
 QY 617 atttaattgcttctgtcgaactcttgaagcttcccaatccagttggaatatgttacaact 676  
 DB 617 acctcaactgcagtgaagacttctccagagctcagtgctgaggaataacttccgataa 676  
 QY 677 tgggttggaagaatttccctgaaggtcaatttgaatttgggtgtcagaggttgaaaaac 736  
 DB 677 catgttggaagcagcttccctataaagaagcagcagagatggtccaggtctccaagatgt 736  
 QY 737 tcaatggaacaatatatgttgaagcaggttgaacccccaacagtaaatgtgaagaattca 796  
 DB 737 tgaatggaggaattagcagtgagctcagaggtatctctcctaagcacaagaaccgctga 796  
 QY 797 cttaaccatgactgaactaagaggtgtccttatgatatgtgaagttaccataaaggaca 856  
 DB 797 aatatactttgagtgagtagagagacataaac-----ctaaca 838  
 QY 857 acatcccaagagagaccccccataacatccaagatttgttcaggtgttatttg 916  
 DB 839 acagaagaagagctcctccctataatattatctatgtttacaggaatgcatatg 898  
 QY 917 tttaagtcagcaattgtttaatatatttacaacaccatcgttcaagacttttg 976  
 DB 899 tggcttcccgagatttgcctcaacatgttttgaagacccttaattcccaacactgattg 958  
 QY 977 ccggttctaagaacacatctcctcagtagagacttggcttaactgtatcgggttc 1036  
 DB 959 aatgggttaagaagcaactttagccagatggaacactctcggccacccctcagctgac 1018  
 QY 1037 caggaatccctggagagatttccagatccagccagatg---tgtctgtcgcagagta 1093  
 DB 1019 gggagatgctcgtgctgttcccaaccaccccaagtagacatctcagacatgacttcta 1078  
 QY 1094 agactcgccttgcagatgaattactatgaagagcttttctacccaatt----- 1143  
 DB 1079 ttgcacagcgtgcgaagtgcagagtcagatgaggaagacatgataaggtgtcctctatg 1138  
 QY 1144 ----tgtactgatactcaactctcgaagcgtgtgtattatgagctgcagaattaggt 1198  
 DB 1139 ctccctctcttggaaatccacagcaggtctatcgtcgtttatgggtgtggtatgaatt 1198  
 QY 1199 ggcattacaagatgagcattgttgcataataattgtatcaggtggaacctatct 1258  
 DB 1199 ggaatgcttcaaacacatcacttggccacaagtttgaaccacaagtagatgataatg 1258  
 QY 1259 tgattaaatgcttggcagaaaagct 1283  
 DB 1259 ctctcagtgcttagaagaataact 1283

DE A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.  
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;  
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;  
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;  
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;  
 KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;  
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;  
 KW diverticulitis; ulcerative colitis; ss.  
 OS Homo sapiens.

XX CA2296936-A1.  
 XX  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 03-FEB-2000; 2000CA-2296936.  
 XX  
 XX 03-FEB-1999; 99US-0118674.  
 XX  
 PA (GLYC-) GLYCODESIGN INC.  
 XX  
 PI Korczak B, Lew A;  
 XX  
 DR WPI: 2000-594746/57.  
 XX  
 PT New nucleic acid molecules of core 2  
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new  
 PT compositions for treatment of disorders mediated by the enzyme  
 PT including cancer, cardiovascular and inflammatory disorders.  
 XX  
 PS Claim 4; Page 51-52; 66pp; English.

XX The present sequence encodes a partial human core 2  
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.  
 CC The polypeptide can be used to treat diseases and disorders, such as  
 CC cancer, cardiovascular disorders and inflammatory disorders including  
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,  
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome  
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as  
 CC atherosclerosis and clotting can also be treated. The polypeptides of  
 CC the invention are predominantly expressed in gastrointestinal tissue  
 CC (stomach, colon, intestine, testis) and are elevated in cancer.  
 CC Gastrointestinal disorders that may be prevented or treated include  
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis  
 CC and ulcerative colitis. The antibodies may be used in  
 CC immuno-histochemical analysis, to detect the novel polypeptide and to  
 CC localize it to particular cells and tissues and to specific subcellular  
 CC locations and to quantitate the level of expression.

XX Sequence 2108 BP: 596 A; 464 C; 509 G; 539 T; 0 other;

Query Match 14.1%; Score 191.8; DB 21; Length 2108;

Best Local Similarity 52.5%; Pred. No. 5.3e-44;

Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtagtgatgacattatcagacccctaagaaggtatgctcaaaagctgtctcaa 376  
 DB 584 tcaccagagactgtagcacttcaagctgaaggaagtcatacagttccaccatgagca 643  
 QY 377 aggaagagaagaagctcccaatagctatcttgggtgtccacaagaagtgaattatg 436  
 DB 644 aagaagaaggtgagctcccatatgcatctatggtgtatcatgagaagattgaanaact 703  
 QY 437 ttgaagagctatccatgatatatacacaagacacatttactcaccattatgac 496  
 DB 704 ttgaagagctacgtcgagctgtgtatgccccctcagaacatactgctccatgtagatg 763  
 QY 497 gtaagacactgataccttcaaaagtgtgcatagacaatttagtaagtctctccacaa 556  
 DB 764 agaagaccctcagaacttccaagaagcggtcaaaagcaatttctctccccaagc 823  
 QY 557 ttctactgttcccaattagagctgtgtgaatgtccacatttcccaaccacagctg 616  
 DB 824 tctctatagccagtaagctcggttcggtgttattgtccctccggtccagggcgcaagctg 883  
 QY 617 atttaattgtctgcgacctctgaagtcctcaatccatgcagtggaatatgttatcaact 676  
 DB 884 acctcaactgcaggaagactgtccagagctcagtgccgttggaatacttccctaata 943  
 QY 677 tgtgtgtgcaagatttccctcgaagtcacatttgaattgtgtcaggttggaanaaac 736  
 DB 944 catgtgtgaggaacttccctcacaagaagcagatgagatgtccaggtctcccaagatgt 1003

QY 737 tcaatgagcaaalatgttggagacgggtgaaccccccaacagtaattggaagattca 796  
 DB 1004 tgaatggaggaataagatgagtcagaggtacactcctaagcacagaagaaaccgctgga 1063  
 QY 797 cttaacctatgaacttagacgggtgccttatgataatgtagagcacaacataagagaca 856  
 DB 1064 aatatcactttagtgatgtgtagagacacataaac-----ctaacca 1105  
 QY 857 acatcccaaggaagcaccocccataacatacagatattgttgcagtgcttatttg 916  
 DB 1106 acaagaagaagatcccccctataatttaactatgatttacaagggaatgcgttaattg 1165  
 QY 917 tttaagtcagagcttgtttaaataatatttcaaacctccatccatgcgttcaagacttttt 976  
 DB 1166 tggcttcccgagatcttgcacaaatgttttgaagaaccttaatcccaacatgattg 1225  
 QY 977 cctgttcaagaacacactctctcgtatgagacacttttggtctactgtatctggatc 1036  
 DB 1226 aatggttaagaagacacttatagccagatgacacacttggccacccctcagctgacac 1285  
 QY 1037 caggaataacttggggagatttccaagatcagcccaagatg---tgtctgacttgcagatg 1093  
 DB 1286 ggtggaatgctgtgctctgtcccaacaccccaagtcagacatccagacatgacttcca 1345  
 QY 1094 agactgccttgcctcaagtgaattactatgaagcttttctatccagct----- 1143  
 DB 1346 ttgcacagctgtgctcaagtggtcaggtatgtagggagacatcgataaggtgtccttatg 1405  
 QY 1144 -----tgtactgtgctccacacttcgaagcgtgtgtattatgtagctgcagattaggt 1198  
 DB 1406 ctccctgtctcggatccacacacagcgggtctatctgcgtttatgggctgggacttgat 1465  
 QY 1199 ggtcttacaagaatggagatgtgttgcataaattgttccaaagtgtagaccctact 1258  
 DB 1466 ggaatgcttcaaaacacacactgtgtgccaagaagttggaacccaaggtatgataatg 1525  
 QY 1259 tgattaaatgcttggcagaaagact 1283  
 DB 1526 ctctcagtgcttagaagaataact 1550

RESULT 5

AAC62134  
 ID AAC62134 standard; DNA: 2128 BP.

XX AAC62134;

XX 06-MAR-2001 (first entry)

XX Nucleotide sequence of beta-1-6-N-acetylglucosaminyltransferase.

XX Human: beta-1-6-N-acetylglucosaminyltransferase; C26NT-M; inflammation;

XX membrane protein; branched sialyl Lex; L-selectin; immune reaction;

XX inflammation; tissue rejection; tumour metastasis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 354..1670

XX /tag= a

XX /product= "beta-1-6-N-acetylglucosaminyltransferase"

XX polyA\_signal 2100..2105

XX US6136580-A.

XX 24-OCT-2000.

XX 19-JAN-1999;

XX 99US-0233506.

XX 19-JAN-1999; 99US-0233506.

PA (BURN-) BURNHAM INST.  
 XX Fukuda M, Yeh J;  
 PI  
 DR WPI, 2001-040238/05.  
 DR P-PSDB; AAB30518.  
 XX  
 PT New C2GNT-M polypeptides having core 2, core 4 and I branching  
 PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing  
 PT reagents useful for diagnosing, preventing or treating inflammation or  
 PT tumour metastasis  
 PS  
 PS Example 1; Fig 4; 25pp; English.  
 XX  
 CC The present sequence encodes a human  
 CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4  
 CC and I branching activities. It is designated C2GNT-M. C2GNT-M is a  
 CC membrane protein that is predominantly expressed in colon, small  
 CC intestine, trachea, stomach and thyroid, as well as in certain cancer  
 CC cell lines. C2GNT-M polypeptides may be used to prepare molecules having  
 CC highly branched steryl hex and L-selectins, which may be subsequently  
 CC used to modulate immune reactions, e.g. inflammation and tissue  
 CC rejection, and to prevent or inhibit tumour metastasis.  
 CC  
 XX  
 SO Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;

Query Match 14.1%; Score 191.8; DB 22; Length 2128;  
 Best Local Similarity 52.5%; Pred. No. 5.3e-44;  
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgatgtgacattatcagactctaaagagttatgctccaaagctgtccca 376  
 Db 670 tcaccagagactgtagacactcaagctgaaggaagtcatacagttccacatgagca 729  
 QY 377 agagggagaaagctcccaataagcctatcttggtgtgccaagaagatgcaattg 436  
 Db 730 aagaagaggtgaggtccctatctgacacactatgagtgatcagaagaatgaaact 769  
 QY 437 ttgaagagctatccatgctatatacaacagacaatattactgcatcattatgac 486  
 Db 790 ttgaagagctactgagactgtgtagtcctccagacaatactgtcctatgtgagtg 849  
 QY 497 gtaagggacccgtacactcaagttgcacatgaacaaattagcgaagtgtccctcata 556  
 Db 850 agaagttcccccgaacttcaagaagcggtcacaagaatattcttgcctcccaatg 909  
 QY 557 ttccatgtcttccaattagaagcgctggaatatgccacattccagaactccagctg 616  
 Db 910 tcttcataagccagaaagctggttgggtttagtctcctcctggtccagagtgcaagctg 969  
 QY 617 atttaattgtcttggaactcttggaagcttcaatccagttggaatattgataact 676  
 Db 970 accccaatgcatggaagactgtctccagagctcagtgccgttgaataactcctgata 1029  
 QY 677 tgtgtgggaagatttccctgaagtcacaatttgaattgtgtcagagttgaaaaac 736  
 Db 1030 catgtggagcagacttctctaagaagcaatgtagcagatgtgtccagctctcaagt 1089  
 QY 737 tcaatggagcaaatatgttgaagcgttgaagccccaagaagaaatttggaaagatca 796  
 Db 1090 tgaatggaggaatagatagagtgccagagtaacctcttaagacaagaagaaacccgctgga 1149  
 QY 797 cttaacatgaacttagagaggtgtccctttagaatttgaagctacaataagaagca 856  
 Db 1150 aatacaacttgagtagtagagacacattaac-----ctaacca 1191  
 QY 857 acatctcaagaagagcaccgcccaataacatcagatattgttggagagtgctatttg 916  
 Db 1192 acaagaagaagatctccctctataatttaactatgttcaaggagaaatgagacattg 1251  
 QY 917 tttaagtcagagcattttaaataatttcaacaactccatcagttcgaagacttttg 976

Db 1252 tgcctcccgagatttcgccaacatgttttgaagaccctaaatcccaacatgattg 1311  
 QY 977 cctgtctaaagaacacactctcctgattgagacttttggcctacacttgattgggttc 1036  
 Db 1312 aatggtaaaagacactatagccagatgaacaccttgggcccacacttcaagctgac 1371  
 QY 1037 caggaataacctgggagatttccagatcacgcccagatg---tgtctgactgcagatga 1093  
 Db 1372 ggtgtgactgtgcctctgttcccaaccccagaagtaagacatcagacatgacttcta 1431  
 QY 1094 agactgcctgtcgaagtgatctatgaagcttttctatccagt----- 1143  
 Db 1432 tgcacaggtctgcaagtgtaggagatgaggaagacatgataaggtgtcctcttg 1491  
 QY 1144 -----tgtactgactcactccctcgaaggtgtgattatgagctcgaataaagt 1198  
 Db 1492 cccctgccttgatccacccagcggtcatctgttattagggcttggagactgatt 1551  
 QY 1199 ggcctatcaagatgagactgttgcataaattgacttaagttgaagttacactatct 1258  
 Db 1552 ggaatgtcaaaacacatccactgttggccaacaagttgacccaaggtatgataatg 1611  
 QY 1259 tgatgaatgcttggcagaagaagct 1283  
 Db 1612 cctctcagtgcttagaagaataact 1636

RESULT 6  
 AAC99109  
 ID AAC99109 standard; cDNA; 2229 BP.  
 XX  
 AC AAC99109;  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.  
 XX  
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KW antiinflammatory; cardiac; gene therapy; chromosome mapping;  
 KW linkage analysis; tissue identification; tissue typing; forensic;  
 KW neural; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055320-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000MO-US05989.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI: 2000-579444/54.  
 DR P-PSDB; AAB54344.  
 XX  
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX  
 PS Claim 1; Page 759-760; 1379pp; English.  
 CC AAC99773 to AAC99731 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, cardiac and antiinflammatory activities, and can be used

CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC9223 to AAC9240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.  
 CC  
 XX  
 SQ Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 14.1%; Score 191.8; DB 21; Length 2229;  
 Best Local Similarity 52.5%; Pred. No. 5.4e-44;  
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgatgtgacattatccagactcaagagtgatctcaaaagctgtctca 376  
 Db 743 taccagagactgtgacactcaagctgaagagtgatctcaaaagctgtctca 802  
 QY 377 agggaggaagaaagctcccaatagctctttgtgtgtccacaagaatgcaatttg 436  
 Db 803 aagaagagtggaagctccctatgcatctctatggtatcaagagaagattgaaact 862  
 QY 437 ttgaagagctatccatgcatatatacaaccagcaaatatttactgtccatgatc 496  
 Db 863 ttgaagagctactgagagctgtgtatgtccctcagaacataactgttccatgtg 922  
 QY 497 gtaagagcccgatcccttcaaaagtgtgcatagaacattgtcgaagtgcttccca 556  
 Db 923 agaagtcgccagaaacttcaaaagagcggtcaagaacattattctgtctccaa 982  
 QY 557 tttaactgtctccaaattagaagcggtgtgaataatgccacattccagctcag 616  
 Db 983 tcttcataagccagtaagctgtgtgtgtgtttatgacctctgtccaggggtgca 1042  
 QY 617 atttaattgtctgtcgaacttctgaagcttcaatccagtggaataatgttatac 676  
 Db 1043 acctcaactgcatggaagactgtctccagagctcagtgccgtggaataactc 1102  
 QY 677 tgtgtgggcaagatttcccccgaagtcgaatttgaattgtgtcagaagttgaa 736  
 Db 1103 catgtgggagcagacttctctataaagaacatgacagatgtgtccagctctca 1162  
 QY 737 tcaatgagcaaatatgtgtgagacggtgaaaccccaacagtaaatgtgaaagat 796  
 Db 1163 tgaatggagagatagcatgagtgatcagtgatccctcaaacacaaacccgt 1222  
 QY 797 cttaaccatcatgaacttgagcgggtgcttgaataatgtgaagctaccataag 856  
 Db 1223 aatataccttgaggtagtgagacacattacac-----ctaacca 1264  
 QY 857 aacatcccaaggaagaccccccataacatcagatatgtgtgagtgatttttg 916  
 Db 1265 acaagaagaagatcctcccccctataatttaactatgtttacaagagatcgta 1324  
 QY 917 tttaagcaagcatttttaatatatttcaacaactcatcgtttaaacttttg 976  
 Db 1325 tggctcccgagatttcgcaacacatgtttgaagaccccaaacccaacactgt 1384  
 QY 977 cctgttctaaagacacatactctctgtatgagacattttgggtcactgtattcg 1036  
 Db 1385 aatggtaaaagacacttatagccagatgaacacaccttggccacccttcagcgt 1444

QY 1037 caggaatacttggggagatttccagatacagccagatg---tgtcatctgcagagta 1093  
 Db 1445 gttgttgctgtgctctgtctcccaaccacagaagacacatctcagacagacttca 1504  
 QY 1094 agactgcttggcaagtggaatttactatgaagcttttctatccagct----- 1143  
 Db 1505 ttgcagagctgtgcaagtggaagtgagagagagagagagagagagagagagagag 1564  
 QY 1144 -----tgtatgattctcccttcgaagcgttgttattatgagctgagattaagt 1198  
 Db 1565 ctccctgcttgaataccaccagcggtctatctgcgttattggtcgtggagcttga 1624  
 QY 1199 gcttatacaagaatggaacattgttgcataaattatgataagtgagagccctatct 1258  
 Db 1625 ggtgtgttcaaaacccatccacttctgtgccaacaagattgaccaagaagtgatgata 1684  
 QY 1259 tgattaatgcttggcagaaagct 1283  
 Db 1685 ctcttcagtgcttgaagaatact 1709

RESULT 7

AAA48623  
 ID AAA48623 standard; CDNA; 2319 BP.

AC AAA48623;

DT 19-SEP-2000 (first entry)

DE Human C2/4GNT CDNA.

XX Human; C2/4GNT; UDP-N-acetylglucosamine; O-glycan biosynthesis;

KW O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 496..1812

FT /tag= a

FT /product= "C2/4GNT"

FT primer\_bind 586..605

FT /\*tag= b

FT primer\_bind 1794..1813

FT /\*tag= c

FT polyA\_signal 2244..2249

FT /\*tag= d

XX WO200034449-A2.

XX 15-JUN-2000.

XX 03-DEC-1999; 99MO-DK00677.

XX 04-DEC-1998; 98DK-0001605.

XX (CLAU/) CLAUSEN H.

XX Clausen H, Schwientek T;

XX WPI: 2000-423407/36.

XX P-PSDB: AAY94492.

XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as

XX probe for the detection of specified glucosaminyltransferase from

XX other species and related organisms

XX Claim 5; Fig 2; 47pp; English.

CC The present sequence encodes human UDP-N-acetylglucosamine:  
 CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase  
 CC (C2/4GNT). The protein is the third member of the family of O-glycan  
 CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4  
 CC based O-glycans on to oligosaccharides, glycoproteins and



```

Db 467 ttgcaaggaacttgaccagagccacatcacagcccttatcctaaggaagac 526
QY 387 aagcttcccaatagcctattcttgggtgtcccaagaatgcaattatggttgaagac 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 tgacttcccttgatataatgcatccatccattgacaccttgaacaccttggaagct 586
QY 447 tatcatgtctataacacagcagcaaatattactgctccatcatatgatacgtaagcacc 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 ctccagggtctattacatgcccacaatatcactggttcatgtgtgataagaaagcaac 646
QY 507 tgatacccttcaaaagtgcacatgaaacaattagtaagtgcttcccaatttcatgtgc 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 aactgaatttaagaatggtgtagagcaacttaagcgtctcccaaacgcttttcgcgc 706
QY 567 ttccaaatagaagcgtgtggaatatgcccacaattccagactccagcgtgtttaaatg 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 ttccaaagttagaacccgtgtgtctatgtagaggtatccaggtccaggtgacctgaactg 766
QY 627 ctgtgcagaccttctgaagcttcaatccagtggaatatgtatcaactgtgtgagca 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 catcagaatcttctgccttcgaggtctcatgtagaagtaagttacacacctgtgagca 826
QY 687 agatttcccttgaaatcaaatittgaattggtgtcagaagttgaaaaaaaacctcaatgtagc 746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 agacttcccttgaaacccaagaagaatagttcagatctgaaagatttaaggttaa 886
QY 747 aatatgtgtgtagaggtggaagaaaccccaacagtaattggaagaattccattaccaca 806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 887 aatatccaccacaggggtgtgtgtcccaagctcaatgcaattgtagcagactaaatagtcca 946
QY 807 tgaacttagacgggtgcttatagaatagtgaagctacacaataagagaacaaactccca 866
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 ccaagtagacccgtgggca-----aagaggttccatgtgatacaacaacagcgttga 1000
QY 867 ggaagaccccccccaataatcagatatgtgtgacggtgttttggtttaagca 926
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1001 accggtcccccacaataatcacaatttaactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1060
QY 927 agcattgttaataatatttcaacaactcactgttcaagacttttgcctgtgtctaa 986
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1061 agagtttgcaacttgtgtctgcatgacccacagcgtgtgtgtgtgtgtgtgtgtgtgtgt 1120
QY 987 agcaaatctctcctgtagagacacttttggtactactgtgtgtgtgtgtgtgtgtgtgtgt 1046
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1121 ggcacacttccagtcctgtagagacttctcggtgtgacccaatagagattccagtggttcc 1180
QY 1047 ttggtagatttccagatcagccagcagatgtgtctgtagctgtagagtaagactgccttgt 1106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1181 ttgcttatgcccacaatgcatcccgactg-----gaaacctcagagctat 1225
QY 1107 caagtgaattactatgaagcgttttctatccagattgtactgagatcccaacttgaag 1166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1226 aaagtgtgagtgaatggaagacagacagagagc---tgccacggtcactatgtatcatg 1282
QY 1167 cgtgttatattatgtagagctgacgaattaaagtgtgcttatcaagaatgagacattgtgttc 1226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1283 tattgtatctatggaagacagagacttaagtggtcgtgtatcccaagcgtgtgttc 1342
QY 1227 taataaattgtattcaagtgtagacctatctgtatcaatgtctgtgcagaaagacctga 1286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1343 taacaagtttagacttaataacccactactactgtgtagatgacctgagagcatg 1402
QY 1287 agaa 1290
   |||
Db 1403 cga 1406

```

RESULT 9  
 AAT16201  
 ID AAT16201 standard; cDNA; 1807 BP.  
 XX  
 AC AAT16201;  
 XX

```

DT 22-APR-1996 (first entry)
DE Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme cDNA.
XX
XX Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; IGT;
KW glycosyltransferase; blood group; I antigen; polylactosaminoglycan;
KW ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 255..1457
FT /tag= a
XX
XX US5484590-A.
XX
XX 16-JAN-1996.
XX
XX 09-SEP-1993; 9305-0118906.
XX
XX 09-SEP-1993; 9305-0118906.
XX
XX (JOL-) LA JOLLA CANCER RES FOUND.
XX
XX Bierhuizen MFA, Fukuda M.
XX
XX MPI; 1996-087019/09.
XX
XX P-PSDB; AAR92474.
XX
XX Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme
PT - used to develop prods. for the study, detection and treatment of
PT pathological conditions involving the enzyme.
XX
XX Example 4; Column 29-32; 29pp; English.
XX
XX A CDNA insert (AAT16201) in pCDNA1-IGT codes for human beta-1,6-N-
CC acetylglucosaminyltransferase, I-branching enzyme (IGT) (AAR92474),
CC an enzyme that converts blood group I antigen to I antigen. The
CC cDNA was isolated from a cDNA library pred. from human PA-1
CC teratocarcinoma cells in vector pCDNA1 following transfection into
CC CHO-Py-leu cells and screening with human anti-I antibodies. The
CC cDNA can be used to regulate the expression of human IGT or to
CC modify its biological function, to produce soluble or membrane-bound
CC forms of IGT in host cells, to breed transgenic animals, and to
CC design antisense oligonucleotides.
XX
XX Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;
XX
XX
XX Query Match 13.6%; Score 185.2; DB 17; Length 1807;
XX Best Local Similarity 52.1%; Pred. No. 3.5e-42;
XX Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```



QY	627	ctcttgagacctcttgaagcttcaatccagctgaagtgaaatgttatcaactctgtgtgggca	686
Db	767	catagagatcttcttgcctctgagcttcacatggaaagtgcgtatccaacctgtgggca	826
QY	687	agatttccccctgaagtcaaatcttgaattgtgtgtcagagcttgaaaaactcaatgagc	746
Db	827	agacttcccccttgaaaacccacaaaggaaatgttcagttctcggaaagattttaaaggta	886
QY	747	aaatcatttgagacggttgaaaccccccaacagtaaatggaagatttcaacttaccatca	806
Db	887	aaatatatccccacagggt	946
QY	807	tgaacttagacgggtgtgctctatgataatgttgaagcttaccacaataagggacaacatccca	866
Db	947	ccaaagagcaccttgggca-----aaagagcttctctatgtgtataaagacaacagcggttga	1000
QY	867	ggaagaccccccccatatacatctagatattgttggcaagtgttattgttttaagta	926
Db	1001	accgctcccccccatatcatctcaacttgaatttgaattgtgtgtgtgtgtgtgtgtgtgt	1060
QY	927	agcatttgttaatatataatttccaacatccatcgttcaagaactttttgtcctgtctaa	986
Db	1061	agagtttggcaacttgttctgtatgataccacagggctgttattgtgtcctcagtggtcca	1120
QY	987	agacacatactctctcatgatagcaacttttgggtctactgattcgtgttccagagatacc	1046
Db	1121	ggacacttcagtcctcgtatgagcatcttcgtgtggacatccatagatgcaggtgttcc	1180
QY	1047	tggggagatttccagatcacccagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1106
Db	1181	tgggtctatgcccacatgcatctccggactg-----gaaactcaaggagcat	1225
QY	1107	caagtggaattcatatgaagcgttttcttccacagttgactgatactcaacctcgaag	1166
Db	1226	aaagtggagtgacatggaagacagacagcagagc---tgccagggccactatgatactg	1282
QY	1167	cgtgtgattttatggagctcgcgaatgaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1226
Db	1283	tatttgtatctatggaaacggagactctaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1342
QY	1227	taataaatttgatctlaagctgagccctatctgtatataatgcttggcagaaaaacttga	1286
Db	1343	taacaagtttgagttatactaccaccoccttactgtggaatgcttgaagactgagcattcg	1402
QY	1287	agaa 1290	
Db	1403	cga 1406	

RESULT 10  
 AAV30006  
 ID AAV30006 standard; cDNA; 1807 BP.  
 AAV30006;  
 AC AAV30006;  
 XX  
 DT 11-AUG-1998 (first entry)  
 DE  
 XX  
 Full length cDNA sequence of human IGT.  
 DE  
 XX  
 Human: beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme;  
 KM IGT; core 2-beta-1,1-N-acetylglucosaminyltransferase; C2GNT;  
 KM treatment; disorder; under expression; type-II hypersensitivity reaction;  
 KM neonatal haemolytic disease; autoimmune haemolytic anaemia;  
 KM thrombocytopenia; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US5766910-A.  
 PD 16-JUN-1998.  
 XX  
 FE 07-JUN-1995; 95US-0488135.  
 XX

Query Match	Best Local Similarity	Matches 502; Conservative	Score 185.2; DB 19; Length 1807;	Indels 24; Gaps 3;
09-SEP-1993; 93US-0118906.	13.6%;	0; Mismatches 438;		
(LJOL-) LA JOLLA CANCER RES. FOUND.	52.1%;			
Bierhuizen MFA, Fukuda M;				
WPI; 1998-361697/31.				
P-PSDB; AAW5662E.				
New nucleic acid sequences and their complementary sequences -				
useful for producing fragment of recombinant human I-branching				
-1,6-N-acetyl-glucosaminyl-transferase polypeptide				
Claim 2; Columns 29-32; 31pp; English.				
The present sequence encodes human				
beta-1,6-N-acetylglucosaminyltransferase, the I-branching enzyme (Ignt).				
The protein sequence has a type II transmembrane topology that consists				
of a short amino terminal cytoplasmic sequence, a signal-anchor sequence				
followed by a short stem region and a large carboxyl terminal catalytic				
domain. The protein shows some homology to core				
2-beta-1,4-N-acetylglucosaminyltransferase (G2Gnt). The nucleic acid				
sequence is used for producing Ignt polypeptides, which can be used				
to treat disorders arising from under expression of Ignt, e.g. increased				
susceptibility to type-II hypersensitivity reactions such as neonatal				
haemolytic disease, autoimmune haemolytic anaemia and thrombocytopenia.				
Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;				
327 ttgtgacattatcagactctaaaggttaagtctaaagaagctgtctcaagaagagagaa	386			
467 ttgcaagaataacttgaccaggcagccactaatacagcccttatctaagaagaagc	526			
387 aagcttcccaatagcctattcttggttgtccacaagaatgaactatggttgaagct	446			
527 tgaacttcccttggaatataatggtatccatcatatcattgacaccttgcagaagct	586			
447 tatcatgctatatataccacagacaacattatctgcatacttaatgatactgaaagcacc	506			
587 cttagcggtattatcagcccccaataatctactacgtgttatactggttgaagaagcacc	646			
507 tgataccttcaaaagtgtccatgaacaatttagctaaagtgcttcccaatatctcatgc	566			
647 aactgaatttaagaatggtgtagagcaactataaagcgtctcccaaacgctttctgc	706			
567 ttccaatttagaggtgtgtggaatagccacaatttccagactccaggtctgatttaaatg	626			
707 ttccaagatgagaccggtgtctatgagggaggtctccaggtccaggtctgacctgaatg	766			
627 ctgtgcgaaccttccgaagttctcaatccaaagtgggaatatgtatcaacttgttgggca	686			
767 catcagaagatcttctgctctcgaaggtctcatggaagaagattatcaaacacctgtggca	826			
687 agatttcccccgaagtcacaatttgaatttgttcagagtgtagaanaaaactcaatgagc	746			
827 agacttccccctgaataacacaagaagaatagttcagatctcgaaagatttaaggttaa	886			
747 aaattgttgtagacgtgtgaaaaccccaaacagttaatttgaagaatctcaattaccatca	806			
887 aaattcccccagaaggtgtctgcgcccgctcatgcaatttgacggaactaaatagttcca	946			
807 tgaatttagacggtgtgcttataatgatatgtgaagttacccaatgaagaacatctccaa	866			
947 ccaagtagaacctgggca-----aagagcttcccatgtgataagaacaaacggttgaa	1000			
867 ggaagacaccccccaataatcagatatatttggcagtgctatttggtttaagcca	926			
1001 accgctctcccccaataatcacaattacttggctctgctatgtggctctctcaag	1060			



QY 1167 cgtgtattatgagcgtcagaaatlaagctggtctarccagaagatgagcattggtttgc 1226  
 DB 1283 tattgtactatggaagaagcacttaagctggtgtaattccaccagcctttggc 1342  
 QY 1227 taataatttgaattcgaagctgagccctatcttgattaaatgcttgcagaaagcttga 1286  
 DB 1343 taagaatttgaagcttaaccacccttaactgagatgctgaactgagagctgcg 1402  
 QY 1287 agaa 1290  
 DB 1403 cga 1406

## RESULT 12

AA061559  
 ID AAC19209 standard; cDNA; 186 BP.

AC AAC19209;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 23284.

KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 9905-0122487.

XX (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS Claim 1; SEQ ID 23284; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

CC Sequence 186 BP; 54 A; 37 C; 38 G; 57 T; 0 other;

QY Query Match

Best Local Similarity 13.4%; Score 183; DB 21; Length 186;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 tttaacctatgctgtctcttcttgaagcttctaagtgcagcactcttcgcga 119  
 |||

DB 1 tttaacctatgctgtctcttcttgaagcttctaagtgcagcactcttcgcga 60  
 QY 120 aaaagacattactggttgcagtaacctccctaagactgcgctttgtaagaacagata 179  
 DB 61 aaaagacattactggttgcagtaacctccctaagactgcgctttgtaagaacagata 120  
 QY 180 cactcatgttaagagcgaagcgaagatgaagttactgttcgggtctcagaacagata 239  
 DB 121 cactcatgttaagagcgaagcgaagatgaagttactgttcgggtctcagaacagata 180  
 QY 240 gcc 242  
 DB 181 gcc 183

## RESULT 13

AA061559  
 ID AA061559 standard; cDNA; 2105 BP.

AC AA061559;

DT 07-OCT-1994 (first entry)

DE cDNA sequence of human core 2 beta 1-6

N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).

KM C2GNT; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;  
 O-glycan; ss.

OS Homo sapiens.

PN

PD

PF

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Location/Qualifiers  
 220..1504  
 /\*tag= a  
 1913..1918  
 /\*tag= b  
 248..314  
 misc\_signal  
 /\*tag= c  
 /label= signal/membrane-anchoring domain

EP590747-A.

06-APR-1994.

29-SEP-1993; 93EP-0250268.

01-OCT-1992; 92US-0955041.

(LJOL-) LA JOLLA CANCER RES FOUND.

Bierhuizen MFA, Fukuda M;

WPI; 1994-111195/14.

P-PSDB; AAR51386.

XX New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are  
 PT used for the study of the effect of variant O-glycan(s) on  
 PT cell-cell interactions, partic. in cancers  
 PS Disclosure; Page 20-22; 34pp; English.

CC C2GNT, or an active fragment thereof, catalyses the formation of  
 CC critical branches in O-glycans. cDNAs encoding various  
 CC glycosyltransferases can be isolated by transient expression of cDNA  
 CC in recipient cells, e.g. COS-1. COS-1 cells were transfected  
 CC with a cDNA library, pCDR alpha-2FL, constructed from poly(A)+ RNA  
 CC of activated T lymphocytes which express the C2GNT. Transfected  
 CC cells were selected using MAb T305, which identifies a  
 CC hexasaccharide on leukosialin. Leukosialin CD43 is an acceptor  
 CC molecule for C2GNT activity. A plasmid, pCDR alpha-Leu,  
 CC which directed expression of the T305 antigen was identified. The  
 CC cDNA insert was isolated and sequenced. The cDNA encoded the  
 CC acceptor molecule leukosialin CD43.

XX Sequence 2105 BP; 614 A; 402 C; 500 G; 589 T; 0 other  
SQ

Query Match	12.9%;	Score 175.2;	DB 15;	Length 2105;
Best Local Similarity	51.9%;	Pred. No. 2.5e-39;		
Matches 517;	Conservative	0;	Mismatches 443;	Indels 36;
				Gaps 4

QY	300	tgaatgattgttgcgaattgcacggtatgttgatgacattttcgaaccttaagggtttagc	359
Db	489	tgaactatataaactgcgcggtacgtgttctctttccatcaaggagcccaatatat	548
QY	360	tcaaaagctgtctcctcaaggaggaagaagctcccaatagccatactctgtgtgcc	419
Db	549	tgtgaagaccctctagttaaaagaagagcggaagttcccaatagatatctatagtgtcca	608
QY	420	caaaagatgcgaattatggtttgaagggtttccatgtcatatatacaccgacaatat	479
Db	609	tcaaaagattggaatgcttgcacaggtctgtgattgccatctatccctcagaatttcta	668
QY	480	ctgcataccattatgactgttaaggcacccatatacctcacaagtgtccatgaaacaattgac	539
Db	669	ttagtctcatgttgagacacaacaacccagagattccctatattgaagctcagttgcatgcgc	728
QY	540	taagtgcttcccaattttcattttctccccaattagaaggtgttggaattggccaat	599
Db	729	tcccggttttaagtaagctctgtgtgcgcagccagatggaggtggttttagaactgtg	788
QY	600	tccagagctccagctgtbatttaattgtctgttcggaacctctcgaagcttccatacagtg	659
Db	789	gagccggtcttcacagctctgacctcactgcgtcatgaaggtctctatgcaatgaaacaactg	848
QY	660	gaatatgatatcaacttgggtgggaagaatttcccccgaagtcaaatlttgatttgt	719
Db	849	gaatctacttatacaactcttgggtgtatgtgatttcccaataaaccacactgaattgt	908
QY	720	gtcagaatttgaaaaaactcaattgagacaataatgttggagaggtgtgaacccccaacag	779
Db	909	caggaagctcaagttgttaatggagaaacaacctggaaccgagagatgcatcccca	968
QY	780	taattggaaagtcttaactaccatcatatgaacttgaagcggtgaccttatgataatgtga	839
Db	969	taaggaagaaaggtcgtgaagaagacgtatga-----gagcgtttaaggaaagctgac	1019
QY	840	gttaccataaagagacaacaactctccagaaggaaagacccccccaataacatcagatatgt	899
Db	1020	-----aaacacagggagctgtcaaaatgtctctccactcgcgaaacacctctcttc	1070
QY	900	tgggaagctattttgttttaagtcgaagcatttgtttaatatatttccaaactccat	959
Db	1071	tgggaagctactctcgtctgtcgtatggatgtatgttgggtttgactacgaatgaagaa	1130
QY	960	catctcaagactcttttgcgcgcgtacataagaacatatctctctgtatgaagcatttggc	1019
Db	1131	aatccaagaattgatggaggtgggcacacaagacacatcacagcctgattgattctctggc	1190
QY	1020	tacttgatccggttccaggaatatactctgggagat-----ttccagatcagcccgagatgt	1076
Db	1191	caccatccaagaagattctcgaagctccggggtcaactccctgcgcagccaatgaatgatct	1250
QY	1077	gtcgtactgcagagtaagagatcgctgttccaagtgaatatactatgaagcgtttt---	1133
Db	1251	atctgacatctgaagcaggttgcgcggttttgtaaggtgcagtaactttgagggtgatttc	1310
QY	1134	-----ctatcccagttgtatcgtgattctcaactctgaaagcgtgtattatarg	1181
Db	1311	caaggtgtctccctcaaccgcgcctgcgaatgagatccatctgctcgaatgtgatttcgg	1370
QY	1182	agctgcgaataaaggtgccttcaaaagatgacatgtgttgcgaataaattgatttc	1241
Db	1371	agctggtgaacttgaacgtgattgtctggcaaacacacactgtttgcataaattgtgaagt	1433
QY	1242	taaggtgacccatcttgaataatgcttggcaga	1277

Db 1431 ggatgttgacctctttgccaaccagtggttgatga 1466

```

RESULT 14
'AAx24042
ID AAx24042 standard; cDNA; 5010 BP

```

AC	AAx24042:
XX	
DT	28-JUN-1999 (first entry)
XA	
DE	Rat DH1 cDNA.
XX	
DH1	rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
KW	diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate
KW	UDP-GlcNAc:galactose-1-3GlcNAc-4-epimerase; transgenic animal; germ line;
XX	beta-1,6-N-acetylglucosaminyl-transferase; ss.

Rattus sp

FH	Key	Location/Qualifiers
000		000

$$/*lag = a$$

XX

XX 1  
C 2  
T 3  
C 4  
C 5  
C 6  
C 7  
C 8  
C 9  
C 10

FD 02 APR 1950.  
XX

PF 02-001-1996; 90CA-2186987;  
XX

PR 02-OCT-1996; 96CA-218698/.

PA (MOON ) MOUNT SINAI HOSPITAL CORP.  
XX

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE  
XX

DR WPI; 1998-399608/35  
DR P-PCND; 1998-399608/35

XX  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  

PT associated with diabetes and hyperglycaemia - comprises reacting

PT in presence of test substance

PS Example 1; Fig 3A; 35pp; English

This invention describes a method for screening for a substance that may be used to prevent or treat cardiomyopathy associated with diabetes and hyperglycaemia. This method involves reacting core 2 GlcNAc-T (UDP-GlcNAc:Galbeta1-3GalNAc-alpha beta1-6-N-acetylglucosaminyl-CC transferase) with an acceptor substrate and a sugar nucleotide donor in the presence of a test substance under conditions whereby the core 2 GlcNAc-T produces a reaction product, determining the amount of reaction product, and comparing the amount of reaction product with the amount obtained in the absence of the test substance. Where lower amounts of reaction product in the presence of the test substance indicates that the substance inhibits core 2 GlcNAc-T. The invention also describes (1) methods for preventing or treating cardiomyopathy associated with diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose germ cells and somatic cells all contain a DNA construct introduced into the animal or an ancestor of the animal at an embryonic stage, where incorporation of the DNA construct into the germ line of the animal causes the animal to develop cardiomyopathy similar to that associated with diabetes mellitus and hyperglycaemia. This sequence encodes the rat DHI protein which is used to describe the method of the invention.

Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match	12.8%;	Score 174.4;	DB 19;	Length 5010
Best Local Similarity	51.8%;	Pred. No. 6e-39;		

Matches 518; Conservative 0; Mismatches 446; Indels 36; Gaps 4;

```

QY 302 atgattgttggaatgacagctgattgtacattatcagactcaagaggtatgctc 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1073 atgactatataaacaatgacccggagctgcgcctcctcattacagacgaacaatatatta 1132
QY 362 aaaaagcttgcctcaaaaggagaaaagctcccaatagcctattcttggttgccaaca 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1133 tggagcccttactaaagaagaggttgccttcccaattgtatccatagttgctcacc 1192
QY 422 aagatgcaattatgattgaaaggcttatccatctatatacaaccgagcaatttact 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1193 ataagattgacatgcttgacagagctcctgagccatctataagcaagaatttctact 1252
QY 482 gcatccattatgactcgttaaggcaccctgataccttcaaatgttgcattgaacaattagcta 541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1253 gcatccacgtgacagaagaagcagaaatccctttagccggcgagcaggtatgcat 1312
QY 542 agtgccttcccaatttctcattgtcttcccaattagaagctgtggaattatgcccacatt 601
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1313 cctgcttgaataagtccttctgtgcccagcttgagagagttgtatccagctcctgga 1372
QY 602 ccagactccagcgtgatttaaatgtctgtcgcaccttctgaagttctcaatccagttga 661
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1373 gtcgggttaagcgtgacctcaactgcatggaagcctgtacagaatgaatgcaacttga 1432
QY 662 aatatgtatcaactgtgtgtggcagaagatttcccttgaaagtcaaatlttgattgtgt 721
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1433 agtactgtacaaactcttctgtgtatgtattccctatlaaacaaccggaattctgta 1492
QY 722 cagaattgaaaacaactcaatgagagcaataatgttgaagacggtgaacccccaagta 781
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1493 ggaagctcaagctcctctacacggggaacacgcttgaaactggaagatgacctcccaaca 1552
QY 782 aattgaaagaatccacttaccatcagaaactgacgggtgaccttgaatagtgaagc 841
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1553 aggaagagaggttgaa-----aaacgatacacggttggagcgaagac 1597
QY 842 tacaataaagaagacaacatctccaaagaaagcaccgcccaatacatcagatatgttg 901
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1598 t-----gaaacaacatgtgattgtcaaggcgagctccacccaactcctcttcttca 1654
QY 902 cagatgtctattcttcttcaagcaagcatttcaatatatttcaacaactcactcg 961
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1655 gcagcgccctattcgtgtcactaggaatatgtatgtatgtcgtggaataaataaata 1714
QY 962 tcaagaacttttgcctgttctaagaacacatacctcctctgtagcacttgggcta 1021
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1715 tcaaaaagttcatgaaatggtggcacaagacacatacagccggatggtcctctgggcca 1774
QY 1022 ccttgattcgggttccaggaataccttggggagatttccagatcagccagagtg--tgt 1078
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1775 ccacccaagaagatcccgaaagtcctgtctcctccccaagccataagtaagtaacttgt 1834
QY 1079 ctgattcggagatgaagactcgcctctgtcaagtggaattactatgaaggttttcc----- 1134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1835 ctgacacgaatgcctgcctcgaagttcgtcaagtggaacacttcgaaagcgagatgttcca 1894
QY 1135 -----tatccaagttgactggaatcacttcgaaagcggttattatagag 1183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1895 atgagcgaccttatccaccgttgagtgagatcattgctcgtctgtgcttggggg 1954
QY 1184 ctgacgaataaagttgctatatacaagaatggaacattggttgcataaatttgattcca 1243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1955 ttggttgcattgagctgabtcttgcgaacaacacacttttgcataaagtgtgacatg 2014
QY 1244 aggtgacacctatctgattaaatgcttggcagaagaact 1283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2015 atgtggtacccttgcctccagtggttggaaagaacatct 2054

```

RESULT 15  
AAC65468

```

ID AAC65468 standard; cDNA: 5010 BP.
XX
AC AAC65468;
XX
DT 12-FEB-2001 (first entry)
XX
DE Diabetic rat heart core 2 GlcNAc-T coding sequence.
XX
KW Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
KW UDP-GlcNAc:Galbeta1-3GalNAcalpha1-6
KW N-acetylglucosaminyltransferase; ss.
XX
OS Rattus sp.
XX
PN US6131578-A.
XX
PD 17-OCT-2000.
XX
PF 02-OCT-1997; 97US-0943058.
XX
PR 02-OCT-1996; 96US-0046876.
XX
PA (KING/) KING G L.
PA (NISH/) NISHIO Y.
PA (KOYA/) KOYA D.
PA (DENN/) DENNIS J W.
PA (WARRE/) WARREN C E.
XX
PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
XX
DR WPI: 2000-678642/66.
XX
DR P-PSDB: AAB30297.
XX
PT Preventing or treating cardiomyopathy associated with diabetes mellitus
PT and hyperglycemia, comprises administering a substance that inhibits
PT core 2 N-acetylglucosamine-T activity.
XX
PS Example 1; Fig 3A; 21pp; English.
XX
CC The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
CC alpha1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.
XX
SQ Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

```

Query Match 12.8%; Score 174.4; DB 21; Length 5010;  
Best Local Similarity 51.8%; Pred. No. 6e-39;  
Matches 518; Conservative 0; Mismatches 446; Indels 36; Gaps 4;

```

QY 302 atgattgttggaatgacagctgattgtacattatcagactcaagaggtatgctc 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1073 atgactatataaacaatgacccggagctgcgcctcctcattacagacgaacaatatatta 1132
QY 362 aaaaagcttgcctcaaaaggagaaaagctcccaatagcctattcttggttgccaaca 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1133 tggagcccttactaaagaagaggttgccttcccaattgtatccatagttgctcacc 1192
QY 422 aagatgcaattatgattgaaaggcttatccatctatatacaaccgagcaatttact 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1193 ataagattgacatgcttgacagagctcctgagccatctataagcaagaatttctact 1252
QY 482 gcatccattatgactcgttaaggcaccctgataccttcaaatgttgcattgaacaattagcta 541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1253 gcatccacgtgacagaagaagcagaaatccctttagccggcgagcaggtatgcat 1312
QY 542 agtgccttcccaatttctcattgtcttcccaattagaagctgtggaattatgcccacatt 601
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1313 cctgcttgaataatgtcttctgtgcccagcttgagagagttgtatccagctcctgga 1372

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 22:43:33 ; Search time 1367.35 Seconds  
(without alignments)  
9415.868 Million cell updates/sec

Title: US-09-645-192-1  
Perfect score: 1362  
Sequence: 1 atgaagatttcaaatgtrta.....atctactaccacatcatga 1362

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	437.8	32.1	821	153	BG434801 602507322
2	408.4	30.0	410	223	AQ005888 CIT-HSP-2
3	393.2	28.9	612	235	AQ900348 HS-215_B
4	350.8	25.8	590	246	A6603057 IM0422E09
5	289	21.2	292	224	AQ091453 HS-3016_B
6	253	18.6	908	220	CNS02A4G
7	221	16.2	973	222	CNS05BAA
8	187	13.7	973	222	CNS05BAA
9	174.2	12.8	1862	192	AK008234
10	170.8	12.5	850	106	AL576150
11	163.2	12.0	970	221	CNS04UIM
12	157.6	11.6	756	220	CNS01UMZ
13	148.8	10.9	526	162	BE005007
14	148.8	10.9	612	165	BE292814
15	139.4	10.2	567	242	A2388491
16	133	9.8	572	165	BE234897
17	133	9.8	635	121	AM842822
18	129.8	9.5	655	24	A1722764
19	127.8	9.4	1050	141	BE867668
20	124.8	9.2	869	106	AL555400
21	121.6	8.9	695	143	BF036111
22	118.2	8.7	731	154	BG468641
23	116.2	8.5	842	141	BE869192
24	115.8	8.5	534	5	AA307800
25	113.8	8.4	671	113	AA307800
26	110.8	8.1	788	21	A1528293
27	107	7.9	428	228	AQ437509
28	102	7.5	348	148	BE863271
29	101.2	7.4	482	141	BF416888
30	99.6	7.3	542	115	AM388627
31	97	7.1	633	16	A1133042
32	96.6	7.1	398	239	A2149281
33	95.8	7.0	622	239	A1747955
34	93.4	6.9	605	239	A2194803
35	92.4	6.8	555	139	BE721246
36	90.6	6.7	443	169	BE755013
37	88.6	6.5	489	139	BE721377
38	88.2	6.5	438	121	AM842550
39	87	6.4	543	144	BF077138
40	86.2	6.3	447	239	A2150403
41	86	6.3	801	153	BG385709
42	84.2	6.2	669	123	AM997414
43	79.8	5.9	574	115	AM388539
44	78.6	5.8	829	220	CNS026YT
45	77.2	5.7	866	153	BG386598

## ALIGNMENTS

RESULT 1  
LOCUS BG434801 821 bp mRNA EST 14-MAR-2001  
DEFINITION 602507322F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4604669 5',  
mRNA sequence.  
ACCESSION BG434801  
VERSION BG434801.1 GI:13341307  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 821)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1340 row: k column: 06  
High quality sequence stop: 709.  
Location/Qualifiers  
1. 821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4604669"  
/clone\_id="NIH\_MGC\_79"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: placenta; Vector: pNDR-LIB (Clontech); Site-1: SfiI (ggccgctggcc); Site-2: SfiI (ggccatagggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTAGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC library."

BASE COUNT 241 a 159 c 163 g 256 t 2 others  
ORIGIN

Query Match 32.1%; Score 437.8; DB 153; Length 821;  
Best Local Similarity 99.5%; Pred. No. 3.2e-107;  
Matches 439; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 922 agccaagcatttgtaataatatttcaaacactcgcctgaacttttgcctg 981  
DB 1 AGTCAAGCATTTGTAATATATTTTCACACTCATCATTCACACTTTTGGCTGG 60  
QY 982 tctaaagacacatctctcgtatgagcactttggtctacacttgatcgggttccagga 1041  
DB 61 TCTAAAGACATACCTCTCTGATGAGCACTTTGGGTACCTTGAATCGGATCCAGGA 120  
QY 1042 ataccctggggagatttccagatcagcccgaggatgtctcgtatcgtgaaggttaagctgc 1101  
DB 121 ATACCTGGGGAGATTTCCAGATCAGCCAGAGTGTCTGATCTGAGATGAGACTCCG 180  
QY 1102 ctgtcagatgggaattactatgaaagcttttctaccagatgtactgacttcacactt 1161  
DB 181 CTGTCTCAAGTGAATTCATTCATGAAAGGCTTTTCTATCCCAAGTGTACTGATCTCACCTT 240  
QY 1162 cgaagcgtgtgtattatgagagctgcagaatttaagttggtttcaagaatgagcattgg 1221  
DB 241 CGAAGCGTGTGTATTATGAGCTGCAGAAATTAAGGGGCTTATCAAGATGACATTTGG 300  
QY 1222 ttgtcataaatgttcttaaggttgagaccctatcttgatlaaagctgtggaagaag 1281  
DB 301 TTTCCTATATAATTTGATTTCTTAAGGTGACCTATCTGATTAAGCTTGGCAAGAAAG 360  
QY 1282 ctgtgaagaacagcagagagacttgactccttgcctcagaagaagtattatgagttaga 1341  
DB 361 CTTGAAAGAAAGCAGAGAGACTGATCATCTTGTCTCAGAAAAAGTTATTTATGATAGA 420  
QY 1342 aatctcactaccacatcatga 1362  
DB 421 AATCTCATTCACATCATCATGA 441

RESULT 2  
LOCUS AQ005888/c 410 bp DNA GSS 27-JUN-1998  
DEFINITION CIT-HSP-2288B17.TF CIT-HSP Homo sapiens genomic clone 2288B17, DNA

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
sequence.	sequence.								
AA005888	AA005888.1	GI:3083333							
SS.									
human.									
Human sapiens									
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:									
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.									
1 (bases 1 to 410)									
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,									
Golden,K., Berry,K., Gaenger,D., Sun,E., Wible,C., Shizuya,H.,									
Simon,M. and Venter,J.C.									
Use of a random BAC End Sequence Database for Sequence-Ready Map									
Building (1998)									
Unpublished (1998)									
Other GSSs: CIT-HSP-2288B17.JR									
Contact: Mary Adams									
Department of Eukaryotic Genomics									
The Institute for Genomic Research									
9712 Medical Center Dr., Rockville, MD 20850, USA									
Tel: 301 838 0200									
Fax: 301 838 0208									
Email: mdadams@tigr.org									
Clones are available from Research Genetics (info@resgen.com). BAC									
end search page:									
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html									
Seq primer: M13-21;									
Class: BAC ends.									
Location/Qualifiers									
1..410									
/organism="Homo sapiens"									
/db_xref="GDB:7149764"									
/db_xref="taxon:9606"									
/clone="2288B17"									
/clone_11b="CIT-HSP"									
/sex="Male"									
/cell_type="Sperm"									
/note="Vector: pBelobAC11; site_1: HindIII; site_2:									
HindIII"									
BASE COUNT	135 a	87 c	81 g	107 t					
ORIGIN									
Query Match	30.0%	Score 408.4;	DB 223;	Length 410;					
Best Local Similarity	99.8%;	Pred. No. 2.2e-99;							
Matches 409; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;					
869 aagacccccccatcacatcagatatattgttgacagtgctatttggtttaagtcacag	928								
410 aagacccccccatcacatcagatatattgttgacagtgctatttggtttaagtcacag	351								
929 cattgttaatatatttcaacaactcagtcgttcaagaactttttgcccgtgctaaag	988								
350 cattgttaatatatttcaacaactcagtcgttcaagaactttttgcccgtgctaaag	291								
989 acacacactctccttgatgacgaacttttggtcctacactgattcgggttcaagaatacctg	1048								
290 acacacactctccttgatgacgaacttttggtcctacactgattcgggttcaagaatacctg	231								
1049 gggagattccagatacgaaccagaatgctgcatctgcagagtaagaactcgcctgttca	1108								
230 gggagattccagatacgaaccagaatgctgcatctgcagagtaagaactcgcctgttca	171								
1109 agtgaatactactgaagagcttttcctatccagctgttactgatatcgaacttcgaagc	1168								
170 agtgaatactactgaagagcttttcctatccagctgttactgatatcgaacttcgaagc	111								
1169 tctgtattatgagagctgcagaatgaagtgcttaccagaatgagacatggttgctca	1228								

Db	50	ATAATTGATCTAAGTGGACCCCTACCTGATTAATGCTTGCGAGAA	1
RESULT	3		
LOCUS	A0900348/c		
DEFINITION	HS-2175_B1.A12.T7 C1T Approved Human Genomic Sperm Library D Homo sapiens genomic clone plate=2175 Col=23 Row=B, DNA sequence.		
ACCESSION	A0900348		
VERSION	A0900348.1	GI:6356538	
KEYWORDS	GS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 612) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		
MEDLINE	99380589		
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC and Web server: <a href="http://www.htsc.washington.edu">http://www.htsc.washington.edu</a> Plate: 2175 row: B column: 23 Seq primer: T7 Class: BAC ends High quality sequence stop: 612.		
FEATURES	Location/Qualifiers		
source	1..612 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate=2175 Col=23 Row=B" /clone_lib="C1T Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBelBAC11; BAC clones in E-Coli DH10B"		
BASE COUNT	180 a 141 c 121 g 165 t 5 others		
ORIGIN			
Query Match	28.9%; Score 393.2; DB 235; Length 612;		
Best Local Similarity	91.6%; Pred. No. 3,1e-95;		
Matches	427; Conservative 0; Mismatches 38; Indels 1; Gaps 1;		
QY	839 agctcaacaataaggacacaacatctccaaggaagcaccocccataacatctcgatatcttg 898		
DB	610 AGCTTCCAATAGAGTCAAGCTTCTCCAAAGGAAGCACCGGCTCATTAACATTCAGATATGTG 551		
QY	899 ttggc-agtgcattattttttaagcaagaatttgttaaatatattttccaacaactcc 957		
DB	550 TTGGCAAGATGATATTTTGTTTCAAGTCAAGCATGTGTTAAATATATTTCAACAACACTCC 491		
QY	958 atcgttcaagaactttttgcctggcttaagaagacatactctccgtatagacattttg 1017		
DB	490 ATGTTCAAAACTTTTGTGGTGCTGTAAAGACACATCTCTCATGATGATGAGCACTTG 431		
QY	1018 gctacattgattcgggttcgaagaatactcgggagatttccagatcagcccaagsgatg 1077		
DB	430 GTTACCTTGATGGGTTCCAGAAATACCGGGGAGATTTCAGATCAGCCCGAGATGTG 371		
QY	1078 tctgattctgcaggttaagatctgctcttgcgaatggaattactatgaagcctttctat 1137		
DB	370 TCTGATCTGCAAGATGAAGACTCGCTTGTCGAATGTGAATTAATCAAGGCTTTTCTAT 311		

Oy	1138	cccgatgtgactgagatctccaccttcgaagctggtgtatcttggagctgcagaattaaag	1197
Db	310	CCCGATTGTACTGGATCTCACCTTCGAAGCCTGTGTATTATTTGAGCGCGCAAGATTAAAG	251
Oy	1198	tggcttatcacaagatggacattggtgtgtcataataatttgatcttaagtgtgaccctatc	1257
Db	250	TGGCTTATCAAAGATGGACATTTGGTTGCTATATAATTGTATTCTAAGGTGGACCATATTC	191
Oy	1258	ttgatataatgctctgcacgaagaagcttgcagaacaacgcagaagaact	1303
Db	190	TTGATTATATGCTTGGCGAAGAGCTTCGCATGCGCGCAGTCGCGCT	145
RESULT	4		
LOCUS	A2603057		
DEFINITION	A2603057	590 bp	DNA
ACCESSION	1M0422E09F	Mouse 10kb plasmid	UGCLM library
VERSION	A2603057	clone	UGCLM0422E09 F, DNA sequence.
KEYWORDS	A2603057.1	GI:11725247	
ORGANISM	GSS.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 590) Dunn,D., Aoyagi,A., Barber,M., Baccorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.		
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts		
JOURNAL	unpublished (2000)		
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112 USR		
FEATURES	Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0422 row: E column: 09 Seq primer: CGTGTAAACGACGCGCCAGT Class: plasmid ends High quality sequence stop: 550. Location/Qualifiers 1..590		

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG1M0422809"  
 /clone\_id="Mouse 10kb plasmid UUCG1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD29v. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/narseq/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD29 (g11473114191AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells

BASE COUNT	and selected for ampicillin resistance."
157 a	147 c
138 g	147 t
1 others	

Query Match	25.8%;	Score 350.8;	DB 246;	Length 590;
Best Local Similarity	79.6%;	Pred. No. 8.4e-84;		
Matches 440;	Conservative 0;	Mismatches 108;	Indels 5;	Gaps 2

QY	1	atgaagatcttcaaatgtaatttattttaaacaataccctcagcaaaagaatttcaatccgttt	60
Db	38	atgaagatattcagatgattgtgcttttaaatcacctccacacagaaactttcaatccctcc	97
QY	61	ttaacccctatgctctctctcttgtttaaagctcttcaaagt--agaagactcttcsg	117
Db	98	tttaaccttctgctgttcttcttcttcttgaagcttcttaaatgttgaggcagacctcttccct	157
QY	118	caaaaagacatttacttggtttgtaactccctaagtaacctcgcttttgaagaaacaga	177
Db	158	caaaagacattttactttacttgaattcctccctaaagtaacatcaaccttttgaggaacag	217
QY	178	tacactcatgttaagaataagcagaagtgaagtaactgttcgggtatctctatgaacag	237
Db	218	tttccccgaattctggggatccagccaaaggacaaacgttttaactgctcggggctctacgacac	277
QY	238	gaagccttggaaatttgaagaaagcttcggaataaagaagaaggaatcatctgaacttggag	297
Db	278	gaagcctttggaaatccggcaagaaacttagaaattcagaagaagcggagctcatctgcattggag	337
QY	298	gatagaagaagtgttgacaaatgaccaatgtttgtgaacttatagaacttaagaactaaagttat	357
Db	338	gaagggaatgtctggcgaatgacaaatgactgtgaactttatagaaccttaagacagatac	397
QY	358	gctcaaaagcttgcctcaaaagagaagaaagcttcccaataagactattcttggtytc	417
Db	398	catgaaaaagctgtttccaaggagaaagaaagaaactttccctatttcgtggtcgtc	457
QY	418	caaaaagatgcaaatgttgtttga--aggttctcatgtgtatatacaaccagacacata	475
Db	458	caaaaagatgcaattttggttggaaggcgggttgatccgagctainttacaacacacacac	517
QY	476	tttaccgtcatcattatgatcgttaagcagcactatataccttcaaagttgcatagaacaatt	535
Db	518	tttaccgtcatcattatgatcgttaagcagcactatataccttcaaagttgcatagaacaatt	577
QY	536	taagtaagtgcct	548
Db	578	taagtaagtgcct	590

RESULT	5
AQ091453/C	
LOCUS	292 bp        DNA               26-AUG-1998
DEFINITION	HS_3016.BL AI_07 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone plate=3016 Col-19 Row=B, DNA sequence.
ACCESSION	AQ091453
VERSION	AQ091453.1    GI:3460364
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 292)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T, Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT:	Contact: Mahairas GG, Wallace JC, Hood L
	High Throughput Sequencing Center
	University Of Washington



OY		361	caaaactgtctcgaagagaagaagaagtccccaatgacctattcttggttgcac	420
Db		448	AAGAGTGGCTGTCGCCAATATGAGACAGACTTTCCCTGGCGTAACATCATGTGTGCAC	389
OY		421	aaagatgccaatcaytggttaaggagcttatcacatgatataacaaccagacaatatcac	480
Db		388	AAATATAGCCTGATGGTGAGAGAGGCTCCTCCAGGGCCGCTGATCCCOCATATACTATTC	329
OY		481	tgatcatcatatgacscgtaagsgaacctgataccctcaagtbcocatgaacaatttagct	540
Db		328	TGCTTCACTAGACCCTGAAGTCCCTCACAGATTCTTCGGCCATAGAAGGGCTGGGCC	269
OY		541	aagtgtcttcacaatatcattcattgcttcacaattagaagcgctvgaaatatgccacatt	600
Db		268	CCTGTGTCGCCAACAGCTGTTCATCCGCTCCAAAGCGGAGGCTGGTTCATTACGGGGCTTC	209
OY		601	tccagactccagcgcgatlltaattgctctfrcggnaccttcgaagcttcaaccagtyg	660
Db		208	AGCCGGCTCAAAGCCGACCTCACTACGCTGTGCGACCTTTTGAGTCAAGAGTCAAGTGG	149
OY		661	aaatacgtatactaactgtgtgyggaagatttcctccctgaagtcacaatttgaattgtyg	720
Db		148	AAATAGCATCATCAACCTCTGCGCGAGATTTTCCCTCCAGTCCAAACATCGAGCTGGTG	89
OY		721	tctgaattgaaaaaacctcaatgagcaaatbtgttgaagscgtytgaaccccacaacagt	780
Db		88	TCGAGACTGAAAGAACTAAGCGGGCTATATGTTAGAGACGGGGCGACCCACATGAGCTAT	29
OY		781	aaatggaagaagctacttac	801
Db		28	AAGAGAGAGATTCACCTTC	8
RESULT		8		
LOCUS		AKO19924		
DEFINITION		AKO19924	1367 bp mRNA HNC 08-FEB-2001	
ACCESSION		AKO19924	Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:533043OR10, full insert sequence.	
VERSION		AKO19924.1	GI:12860326	
KEYWORDS			CAP TRAPPER.	
SOURCE			Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to mRNA, clone_1lb:RIKEN full-length enriched mouse cDNA library clone:533043OR10.	
ORGANISM			Mus musculus	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.	
REFERENCE			1 (sites)	
REFERENCE			Carninci,P. and Hayashizaki,Y	
REFERENCE			High-efficiency full-length cDNA cloning	
REFERENCE			Methods Enzymol. 303, 19-44 (1999)	
REFERENCE			2 (sites)	
REFERENCE			Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
REFERENCE			Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
REFERENCE			Genome Res. 10 (10), 1617-1630 (2000)	
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
4 (sites)				
The RIKEN Genome Exploration Research Group Phase II Team and				









BASE COUNT 262 a 221 c 245 g 213 t 29 others  
 ORIGIN

Query Match 12.0%; Score 163.2; DB 221; Length 970;  
 Best Local Similarity 56.1%; Pred. No. 3,7e-33;  
 Matches 291; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

204 gtataaatttaacgtcttcgggtatctatgaacaggaagcccttggaaattggaaagatct 263  
 Db 519 GTACAGCTTATATGTGCTCCCGATATAGCAATGGAACCGATGAGAGTGGGAGAGTCTT 460  
 Qy 264 ggaataaagaagaaggaacatcattgacttggaggatgatgatctgttggcaatggaccag 323  
 Db 459 GNTATCCGGAGGAAACCTCTGCCCGAGTCCGACACAAAGCCCTGACCAAACTGACCTN 400  
 Qy 324 tgattgacatttatacagacttaagaggtatgctcaaaagcttgcctcaagaggaga 383  
 Db 399 TGACTGCAAGAGCTTGTTCAGGGCCAGAGTTCAGACAAAGAGTGCNNTTGGAAATGGA 340  
 Qy 384 gaaaagctcccaatagcattcttctgtgtcccaagaagatgcaattatgttgaag 443  
 Db 339 GCAGAGANTTTCCCTCGGGTACTCAATGTGTGCACANATATGCTGATGTGTGAGAG 280  
 Qy 444 gcttatacgtctatatacacaacagacaatatttactgcatcatttgaatgtaagc 503  
 Db 279 GCTCTCAGGGCCGTGTACTCCCAATATACATCTACTGCTCTCCACTGACCTGAACTC 220  
 Qy 504 acctatataccttcaaaagtctgcaatgaacaattgaagtgtcctccaaatttctat 563  
 Db 219 CCCCTACACAGTTCATNTCCGGCCATAGAGGGCCCTGCTGCTGCTGCCAACNTCTCAT 160  
 Qy 564 tgcctcaaaatgaaggcgtgtgaaatagccacatctccagaccagcggctgatttaa 623  
 Db 159 CGCGTCCAAAGCGGGAGGTGTTCATACGGGGCTTCACCGCGCTGAAAGCCGACTTAA 100  
 Qy 624 ttgctgtcgaagccttctgaagcttcaatcagagtggaatattgtatacaacttgtg 683  
 Db 99 CTGCGCTGTGAGCTCTGTAGTCAAGAGTCAAGTGAAGTACGATCATCAACTCTGCG 40  
 Qy 684 gaaagatttccccgaagaatcaatttgaattgtgtc 722  
 Db 39 CCAGATTTTCCCTCAGGTCACATCGNNNTGNTNC 1

RESULT 12  
 CDS101W2 756 bp DNA GSS 12-MAY-2000  
 LOCUS Tetraodon nigroviridis genome survey sequence PUC-01 end of clone  
 DEFINITION 19705 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 ACCESSION AL168380  
 VERSION GI:7806437  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 REFERENCE 1 (bases 1 to 756)  
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizes, C., Fisher, C.,  
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
 Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 Unpublished  
 JOURNAL 2 (bases 1 to 756)  
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizes, C., Winkler, P., Brothier, P., Quetier, F.,  
 Saurin, W. and Weissenbach, J.  
 Human gene number estimate provided by genome wide analysis using

JOURNAL Tetraodon nigroviridis DNA sequence  
 REFERENCE Unpublished  
 3 (bases 1 to 756)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases  
 COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES  
 source  
 1..756  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="19705"  
 /clone\_1ib="G"  
 /note="Genoscope sequence ID : C0A6197CC03SP1-end :  
 PUC-01"

BASE COUNT 169 a 189 c 224 g 166 t 8 others  
 ORIGIN

Query Match 11.6%; Score 157.6; DB 220; Length 756;  
 Best Local Similarity 61.2%; Pred. No. 1.1e-31;  
 Matches 267; Conservative 2; Mismatches 166; Indels 1; Gaps 1;

Qy 848 taagagacaacatctcgaaggaagcaccacccataacattcagatatgttggcagtg 907  
 Db 754 TCMAACACAGCTGCGCAAGCCCGCCCGCCGCGCATCCAGKCTTCATGGCAGCG 695  
 Qy 908 ctatttgtttaaagtaagcatttggtaataatatttcaaacactcatcgttcaag 967  
 Db 694 CTATATTCGTCYTGTCGCGGAGCTTTTGGCACACGGAACGAGAGGAGCTGCCAGG 635  
 Qy 968 acttlttgcgtgtcctaaagacacacacacacacacacacacacacacacacacacac 1027  
 Db 634 ACTTCCTGCGCTGTGTGCTGACCTGACACCTTACCTGCGGAGCAGCTTGTGGCCACCTCG 575  
 Qy 1028 ttccggttccagaataccttggggagatttccagatcaagccagagatgtctgactgc 1087  
 Db 574 TCAGGTCGCCGGGGGCTGCCGCGCACATGCCCGCTCCCAAGCGAGTGCACGATCTGA 515  
 Qy 1088 aagatgaagcctgcgttgaagtggaattactatgaagcctttctatcccaagttgta 1147  
 Db 514 GAAGTAAGACCGCGCTGTGTAATGAACTACTGTGAGGGAGAACTGTACCCGCGCTGCA 455  
 Qy 1148 ctgatactcacctcgaagcgtgtgatttataaggagctgagaattaaagtggttata 1207  
 Db 454 CGGGCACACACCTGCGGAGCGGTGCAATACGCGCGCGGAGCTT-CGCTGGCTGCTCG 396  
 Qy 1208 aagatgaagcattgttgcctataatattgattttaaagtggaacccattgattaat 1267  
 Db 395 GCTTCGACACTGCTGTGCGCAACAAGTTTACCCCAAGTGAAGACCGGCTTCTGATCACT 336  
 Qy 1268 gcttggcagaagaagct 1283  
 Db 335 GTTGGAGAGAGAGCT 320

RESULT 13  
 BE005007/c 526 bp mRNA EST 05-JUN-2000  
 LOCUS MRO-BN0115-020300-001-a03 BN0115 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE005007  
 ACCESSION BE005007  
 VERSION BE005007.1 GI:8265240  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 526)  
 Dias Neto, E., Garcia Correa, R., Veijovski, Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

## TITLE

Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.  
Shogun sequencing of the human transcriptome with ORF expressed sequence tags

## JOURNAL MEDLINE

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663

## COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=st2-MR0-BN0115-020300-001-a03&ts=2000-03-02&ft=1)

Seq primer: puc 18 forward

High quality sequence start: 17

High quality sequence stop: 526.

## FEATURES

## source

Location/Qualifiers

1..526

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0115"

/dev\_stage="Adult"

/note="Organ: breast, normal; Vector: puc18; Site: 1; Sma1; Site: 2; Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## BASE COUNT

126 a 125 c 136 g 139 t

## ORIGIN

## Query Match

Best Local Similarity 10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Matches 246; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Matches 246; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Matches 246; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Matches 246; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

Matches 246; Conservative 0;

10.9%; Score 148.8; DB 162; Length 526;  
Pred. No. 2.4e-29; Mismatches 162; Indels 0; Gaps 0;

RESULT 14

BE292814

LOCUS

601105179P1 NIH\_MGC15 Homo sapiens cDNA clone IMAGE:2988090 5', mRNA sequence.

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..612

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:2988090"

/clone\_lib="NIH MGC.15"

/tissue\_type="adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pOTB; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT

181 a 137 c 140 g 154 t

ORIGIN

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Matches 340; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Matches 340; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Matches 340; Conservative 0;

ORIGIN

10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Query Match

Best Local Similarity 10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

Matches 340; Conservative 0;

10.9%; Score 148.8; DB 165; Length 612;  
Pred. No. 2.5e-29; Mismatches 272; Indels 18; Gaps 1;

```

QY 755 tggagagcgtgaaaccccccaacagtaattgtgaagaattcacttaacatcagaactta 814
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 TGGAGTCAGAGGTACCTCCCTTAAGCACAAGAAACCCGCTGGAAATATACATTGAGGTA 420
QY 815 gacgggtgccttatgaatatgtgaagtaacaaataagacaacatctccaaggaagcac 874
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TGAGAG-----ACACATTACACCTAACCAACAGAAAGAGAGATCTTC 462
QY 875 ccccccaacattcagatatgtgtggcagtgcttatattgttttaagcaagcatttg 934
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 CCCCTTATATTTAATTAATGTTTACAGGGAATGCGTATGCGCTTCCCGAGATTGCG 522
QY 935 ttaataattttcaacaactcgcgtcaagaacttttgcctgtcctaagacacat 994
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 TCAACATGTTTGAAGAACCCCTTAATCCGACACACTGATGTAATGGTTAAAGACACTT 582
QY 995 attctcctgagcagcatttggggtacct 1024
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 583 ATAGCCAGATGAACACCTCTGGGCCACT 612

```

```

RESULT 15
A2388491 567 bp DNA GSS 02-OCT-2000
LOCUS A2388491 567 bp DNA GSS 02-OCT-2000
DEFINITION Mus musculus genomic
ACCESSION A2388491
VERSION A2388491
KEYWORDS GSS.
SOURCE house mouse
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 567)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,
and Wright, D., Weiss, R.

```

```

TITLE Mouse whole genome scaffolding with paired end reads from 10Kb
Plasmid inserts
Unpublished (2000)

```

```

JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

```

```

Insert Length: 10000 Std Error: 0.00
Plate: 0148 row: G column: 03
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 567.

```

```

FEATURES
source
1. 567
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U081M0148G03"
/clone_lib="Mouse 10Kb plasmid U081M library"
/sex="Male"

```

```

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The

```

adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (911473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 124 a 138 c 143 g 162 t  
 ORIGIN

```

Query Match 10.2% Score 139.4; DB 242; Length 567;
Best Local Similarity 54.9%; Pred. No. 8.4e-27;
Matches 275; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 316 atgacacgtgattgtgaattatcagaactcgaaggtatgtctcaaaagctgtctca 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 561 ATGACAGCAGACTGTGAGACTTCAAGACCAAGAGAGATTATACAGTCCCACTGAGC 502
QY 376 aaggaggaagaagcttcccaatagctattcttgggtgtccaaagaagcattatg 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 AAGGAGAGAGGCCAGCTTCCCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 442
QY 436 gtgaaaggtatcgaattatcagaactcgaaggtatgtctcaaaagctgtctca 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 441 TTGGAAGGTGCTGCGAGCTGTGACACCCCTGAGATGATGCTGCTGACATGAT 382
QY 496 cgttaagcaccgtgatacttcaaaagttccatgaacaattagctaaagcttccaat 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 381 CAGAGGTCTTCGAAACCTTTTAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 322
QY 556 atttcattgcttccaaatgaaggtgtgtgaatatgcccaatttccagaactcaagct 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 GCTCTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 262
QY 616 gatttaaatgtgtgtggaacttctgaagcttcaatcccaagtggaatattgtatcaac 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 GACCTAAACCTGATGGAAGACTTGTTCAGAGCCCGCGCATGGAATACCTCTGTAAC 202
QY 676 ttgtgtggaagatttcccccgaagcatttgaattgtgaagtggtcagagttgaaaaa 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 201 ACTGTGGAGACACTTTCCTCAAAACCAATGCTGAGATGCTGAGAGCCCTCAAGCTA 142
QY 736 ctcaatgagcaaatatgttggagacggttgaaccccccaacagtaaatgtgaagaatgc 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 141 TTGAAGAGGCAAGACAGATGAGAGTACAGAGTACACCTCCACATTAATAATCCCGTGG 82
QY 796 acttaccatcatgaactaga 816
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 81 AATATCTACTATGAGGTGACA 61

```

Search completed: August 3, 2001, 00:31:19  
 Job time: 6466 sec

Fri Aug 3 09:38:07 2001

us-09-645-092-1.rst

---

Page 14